This is a digital copy of a book that was preserved for generations on library shelves before it was carefully scanned by Google as part of a project to make the world's books discoverable online.

It has survived long enough for the copyright to expire and the book to enter the public domain. A public domain book is one that was never subject to copyright or whose legal copyright term has expired. Whether a book is in the public domain may vary country to country. Public domain books are our gateways to the past, representing a wealth of history, culture and knowledge that's often difficult to discover.

Marks, notations and other marginalia present in the original volume will appear in this file - a reminder of this book's long journey from the publisher to a library and finally to you.

Usage guidelines

Google is proud to partner with libraries to digitize public domain materials and make them widely accessible. Public domain books belong to the public and we are merely their custodians. Nevertheless, this work is expensive, so in order to keep providing this resource, we have taken steps to prevent abuse by commercial parties, including placing technical restrictions on automated querying.

We also ask that you:

+ *Make non-commercial use of the files* We designed Google Book Search for use by individuals, and we request that you use these files for personal, non-commercial purposes.

+ *Refrain from automated querying* Do not send automated queries of any sort to Google's system: If you are conducting research on machine translation, optical character recognition or other areas where access to a large amount of text is helpful, please contact us. We encourage the use of public domain materials for these purposes and may be able to help.

+ *Maintain attribution* The Google "watermark" you see on each file is essential for informing people about this project and helping them find additional materials through Google Book Search. Please do not remove it.

+ *Keep it legal* Whatever your use, remember that you are responsible for ensuring that what you are doing is legal. Do not assume that just because we believe a book is in the public domain for users in the United States, that the work is also in the public domain for users in other countries. Whether a book is still in copyright varies from country to country, and we can't offer guidance on whether any specific use of any specific book is allowed. Please do not assume that a book's appearance in Google Book Search means it can be used in any manner anywhere in the world. Copyright infringement liability can be quite severe.

About Google Book Search

Google's mission is to organize the world's information and to make it universally accessible and useful. Google Book Search helps readers discover the world's books while helping authors and publishers reach new audiences. You can search through the full text of this book on the web at `http://books.google.com/`

RIVERSIDE TEXTBOOKS
IN EDUCATION

EDITED BY ELLWOOD P. CUBBERLEY
PROFESSOR OF EDUCATION
LELAND STANFORD JUNIOR UNIVERSITY

DIVISION OF SECONDARY EDUCATION
UNDER THE EDITORIAL DIRECTION
OF ALEXANDER INGLIS
ASSISTANT PROFESSOR OF EDUCATION
HARVARD UNIVERSITY

STATISTICAL METHODS APPLIED TO EDUCATION

A TEXTBOOK FOR STUDENTS OF EDUCATION IN THE QUANTITATIVE STUDY OF SCHOOL PROBLEMS

BY

HAROLD O. RUGG
ASSISTANT PROFESSOR OF EDUCATION
THE UNIVERSITY OF CHICAGO

HOUGHTON MIFFLIN COMPANY
BOSTON NEW YORK CHICAGO
The Riverside Press Cambridge

The Riverside Press
CAMBRIDGE . MASSACHUSETTS
U . S . A

EDITOR'S INTRODUCTION

Such a volume as the present number in this series of textbooks forms an interesting exhibit of the progress at present being made in the organization of instruction in the subject of education. Two decades ago there would have been almost no use for such a volume, as we had not then begun to make any accurate measures of the products of our educational efforts. Only the most general terms were then in use, while to-day the demand is for quantitative expression in commonly used terms which students can understand. Especially within the past decade has there been a remarkable evolution of standards for educational work and quantitative units of measurement. To-day the educational investigator and the superintendent of instruction alike need to use refined tools in the measurement of educational results. To such, and to the students in our schools of education generally, the simple presentation of the mathematics underlying the accurate measurement and plotting of educational results here presented should prove of large usefulness.

The author of this volume has stated the aims and purposes and plan of the work so well in his preface that little remains that an editor needs to say. The volume represents a very successful attempt to produce a book which will apply the mathematical theory of statistical work to educational problems, and as such it should find a hearty welcome from teachers of education in universities, colleges, and normal schools, educational investigators generally, and school officers interested in making the best use of statistical data and displaying the results to their supporting public in the most effective graphic form. The author has been particularly fortunate in the selection of what to include in the volume, and in the organization and presentation of what he has included.

ELLWOOD P. CUBBERLEY

PREFACE

During the past two decades a body of quantitative technique has developed in education which makes constant use of technical statistical methods. The school man, in trying to keep pace with the developing tools, has constantly demanded a complete exposition of them. At the same time he has made it very clear that the treatment which will appeal most pertinently to his needs must be couched in non-mathematical language. He has said frankly that his mathematical training has been limited to high-school algebra, and rather an ancient and, in some sense, obsolete algebra at that. He has told us that "graphs" are mysterious things to him; that equations of lines and formulæ have no significance; that the use of "frequency distributions," "probability curves," "medians," "measures of variability," and "coefficients of correlation" can hardly be said to lend clearness to his thinking about his own school problems.

Three courses are open to the writer who wishes to acquaint such persons with statistical methods of treating facts. First, he can say that the school man's lack of familiarity with college algebra, analytic geometry, the calculus, and least squares is his own lookout, and that it is impossible to write a "statistical methods" and to give statistical training without presupposing this particular kind of equipment. We have available now several books and many monographs built on that basis which make use, more or less in detail, of the higher mathematics, but none of which are applied to educational problems.

Second, he can give the student of education a manual of formulæ and rule-of-thumb methods of computing the various coefficients, without any explanation of the derivation of these constants, without an adequate exposition of how to discriminate the use of the different methods, and without making possible a complete and proper interpretation of the results of using the methods. To do this would commit the writer to the rather current theory that, for the educationist, "statistics is arithmetic," and that his statistical equipment should include only the ability to compute the various coefficients and to follow rule-of-thumb methods of interpreting them (*e.g.*, the rule that a coefficient of correlation of say .25 is "high," "low," or "what-not"). The few books and chapters of books which have so far applied statistical methods to school problems have been very largely committed to this doctrine.

Third, the writer in this field can assume that it is necessary to equip school men, generally, with a thorough-going knowledge of statistical methods; that in order for them to be discriminating in the use of the various methods in improving their school practice, this large background of knowledge must be developed; and that it is possible to explain rather completely the reasons for and the significance of the principal statistical devices without expressing the explanation in technical mathematical language.

This book has been written with a deep-rooted conviction that the third of these three courses is the proper one; with a complete recognition of the limitations in mathematical equipment of the "average" school administrator and teacher, which is the outcome of considerable classroom contact with this particular kind of student. It is based upon the knowledge, however, that it is possible to make clear the significance and proper use of the more important statistical devices without expressing these in mathematical form.

The necessary substitution of words for symbols in the explanation of the derivation and common-sense significance of such devices has resulted in what, to the mathematically trained reader, will seem to be a "wordy" book. The prerogative of the "author's preface" leads the present writer to say frankly that in this book he has not been interested in writing for the mathematically equipped reader. At the same time, it is hoped that such a person can, indeed, get an initial view of statistical methods from the following chapters which he can use to advantage in a study of the secondary and original works of Yule, Bowley, Elderton, Karl Pearson, and others who have constructed our statistical tools.

The book throughout has been written in intimate contact with graduate classes in education. It is the direct outgrowth of mimeographed notes written for seven of such classes, and elaborated and revised distinctly in terms of their specific needs and interests. Symbolic and word explanations have given way to graphic devices wherever necessary and possible. The many repetitional "back references," restatements of principles, reasons, etc., in succeeding chapters have been made with a full recognition of the possible inelegance in form, but with a firm conviction in the value of the resulting increase in clearness to the reader. Traditional usage in the form of textbook writing has been deliberately sacrificed to the one criterion of readableness.

A very small group of students of education have made use, recently, of certain methods which have not been included in the discussion of this book. Outstanding among these is Yule's *Partial Correlation*, and Spearman's methods of "correcting" coefficients of correlation. To a very small group of educational psychologists these may seem unpardonable omissions. However, neither set of methods could have been presented in the complete fashion necessary in the treatment of those topics without encroaching unduly upon the limited space of this textbook, already devoted to more important methods. Furthermore, it is doubtful if the former of these methods will be used by more than a very small fraction of those working in educational research in our own generation. These persons should turn to Yule's complete original discussion. In regard to the latter of the two sets of methods, the writer is one of those who are still skeptical of the use of methods of "correcting" coefficients (the validity of which has not been established) which have been computed from material collected under conditions subject to such gross inaccuracies as are the conditions of educational research.

It is fundamental to a clear comprehension of the writer's point of view to know that this book is based upon the doctrine that statistical methods in themselves prove nothing, — that the methods selected for use in a particular situation must agree with the logic of that situation: in a word, that statistical methods are merely quantitative devices which we can use to refine our thinking about complex masses of data, and to refine our methods of expression.

The example of Leonard P. Ayres in his discriminating use of statistical methods in school research, and his constant subordination of the exhibition of statistical form to clearness and simplicity of presentation, has been a potent factor in determining the writer's point of view, and has wrought a definite effect upon his practice.

HAROLD O. RUGG.

SCHOOL OF EDUCATION,
UNIVERSITY OF CHICAGO,
August 22, 1917.

CONTENTS

LIST OF DIAGRAMS

Nov 13.

LIST OF TABLES

STATISTICAL METHODS APPLIED TO EDUCATION

CHAPTER I

THE USE OF STATISTICAL METHODS IN EDUCATION

PROBLEMS AND METHODS IN SCHOOL RESEARCH

Steps in the development of "scientific education." There are two groups of persons in the educational world who are directly interested in the application of statistical methods to school problems — the school administrator, and the teacher and educational psychologist. In correspondence to these two classes of interest, school problems may be said to be either administrative or pedagogical-experimental in character. They arise either in connection with the attempt of the administrative agents of a school system to fit the "machinery of the system" to the needs and capacities of children, or to the attempt of the school man and the psychologist, working together, to determine more minutely the status of learning in the child. The school man's chief concern, then, is with these questions: First, how does the child learn? Second, how may the course of study, methods of teaching, modes of classifying and promoting children, methods of organizing the school year, safe-guarding the health of school children, etc., be best adapted to the established facts of development and processes of learning in children, and to the needs of their future life.

The method of attacking the solution of such fundamental questions prior to our own generation was clearly traditional and based on individual experience. It was said by the representatives of the established sciences, and freely admitted by the pedagogues, that "education" was not a "science" — that its method was not "scientific." By this was meant that school men did not make use of the fundamental steps in the scientific procedure of solving problems.

Fundamental steps neglected. To be specific: (1) They did not systematically observe educational conditions, or collect necessary facts, recording their observations minutely. More concretely this meant that they did not set about collecting the facts on the composition, training, certification, tenure, pay, and rating of the teaching staff; the content of courses of study; the age-grade distribution of pupils, and their progress through the grades; the corresponding measurement of instruction and the capacities of pupils; the extent of their elimination from and retardation in the public schools; the many facts concerning the central administration and organization of schools; school costs, school accounting, and the efficiency of business management; operation of the plant and the handling of equipment and supplies, — the determination of each of which is necessary to the promotion of efficient school administration. Thus, the first step in the utilization of the scientific method — the collection of large numbers of facts — was not taken.

(2) The indictment of our traditional pedagogy pointed out that students of "pedagogy" did not "measure" the results of school work, that no yardsticks were available by which the efficiency of school administration or school teaching could be evaluated; hence that little progress in the improvement of either one could come about. Nobody knew accurately to what extent boys and girls had mastered the elements of reading, writing, arithmetic, spelling,

geography, history, and language. We simply knew that there was an accumulation of incapacity in particular grades of the public schools, relieved in part by rapid elimination of pupils from school.

(3) In pedagogy, however, it was evident that since almost no collection of facts was made that no recourse was had to the development of mathematical or statistical methods of treating the data. Large quantities of data accumulating in the biological and physical sciences had demanded and led to the development of sound methods of statistical treatment in those fields. Prior to fifteen years ago "pedagogy," however, had made no use of the large body of statistical technique that had been put together.

(4) "Science" demands as the capstone of its procedure the utilization of a thoroughgoing experimental attack on the problem in question. Conditions must be "controlled" by the investigator, measurements must be made as minutely as possible, records of results must be kept, and the data which have been collected must be systematically organized through the utilization of valid statistical methods. Again, prior to our generation, this experimental procedure had not been used in education. It is true that four decades ago various German psychologists began the study of "learning" under isolated conditions, and with fairly refined laboratory technique planned a way for the transfer of their technique and certain of their grosser conclusions to classroom analysis of learning and teaching. This actual transfer, however, has come about in our own time.

Lack of thorough collection of facts concerning educational conditions, measurement of results, statistical treatment of the data, setting up of experimental methods of studying school practice, — these are the counts on which the older "pedagogy" was indicted.

Recent developments. The above statement of the ways in which pedagogy failed to utilize scientific method reveals specifically the steps in the development of "scientific education" during the past two decades, and leads naturally to an exposition of the use of statistical methods to school men. The school man has turned to exactly these steps of procedure in the attempt to determine the present status of school practice and to direct scientifically the course of its development.

We have said above that school problems were either administrative or pedagogical-experimental in character. Our first task in taking up the study of "statistical methods applied to educational problems" is to recognize clearly the various school problems whose solution demands treatment by numerical methods. During the past fifteen years every phase of school administration and pedagogy has been subjected to quantitative methods of study and experimentation. Our educational literature abounds with "factual" studies, our educational conventions are given over very largely to discussions of "measurement" and "standardization" of school processes. Outstanding at the present time, therefore, is the need for a clear, scientific, and complete statement of the statistical and graphic methods which the school man must call to his aid in this quantitative attack on educational problems.

To get sharply before us a picture of the new emphasis, let us turn briefly to a few examples of the use of statistical methods in education. These have been so selected that the general field will be brought in review.

I. QUOTATIONS FROM RECENT QUANTITATIVE LITERATURE

1. Measuring reading ability

The checking-up of the results of school teaching by standardized tests is one of the most promising phases of the new movement. The tabulation and classification of the results of testing has led to the development of devices for recording school facts and for presenting the data. The need for tabular and statistical methods is well illustrated by the following quotation from Brown.[1]

TABLE 1. GRADE RECORD

Form A

Town X — *Date of Test, June 4, 1915*
School A — *Grade III*

Pupil No.	*Name*	*Age*		*Rate of Reading*	*Deviation*	*Comprehension*	*Deviation*	*Reading Efficiency*	*Deviation*
		Yr.	*Mo.*						
1				1.28	−1.79	75	+35	96.00	− 18.79
2				1.60	−1 47	38	− 2	60.80	− 53.99
3				1.85	−1.22	63	+23	116.55	+ 1.76
4				2.07	−1.00	50	+10	103.50	− 11.29
5				2.15	− .92	40	....	86.00	− 28.79
6				2.33	− .74	70	+30	163.10	+ 48.31
7				2.36	− .71	33	− 7	77.38	− 36.91
8				2.38	− .69	58	+18	138.04	+ 23.25
9				2.38	− .69	50	+10	119.00	+ 4.21
10				2.63	− .44	36	− 4	94.68	− 20.11
11				2.98	− .09	44	+ 4	131.12	+ 16.33
12				2.98	− .09	22	−18	65.56	− 49.23
13				3.00	− .07	17	−23	51.00	− 63.79
14				3.15	+ .08	22	−18	69.30	− 45.49
15				3.26	+ .19	11	−29	35.86	− 78.93
16				3.28	+ .21	33	− 7	108.24	− 6.55
17				3.32	+ .25	55	+15	182.60	+ 67.81
18				3.78	+ .71	32	− 8	120.96	+ 6.17
19				4.30	+1.23	32	− 8	137.60	+ 22.81
20				4.60	+1.53	29	−11	133.40	+ 18.61
21				5.83	+2.76	60	+20	349.80	+235.01
22				6.02	+2.05	14	−26	84.28	− 30.51
	Average			3.07	0.90	40	15	114.79	40.39

DIAGNOSIS OF CLASS AND INDIVIDUAL NEEDS

In Table 1 are given, for purposes of illustration, the data from an actual third grade. This grade stood second among thirteen

---

[1] Brown, H. A. *The Measurement of the Ability to Read.* Bulletin no. 1, Bureau of Research, New Hampshire Department of Public Instruction, Concord, N.H.

third grades which were tested, and represents a somewhat satisfactory efficiency. Examination of the averages shows that the *rate*[1] of reading, the *comprehension*, and the *reading ability* of the class as a whole are high. The *rate* of reading is seen to be very high.

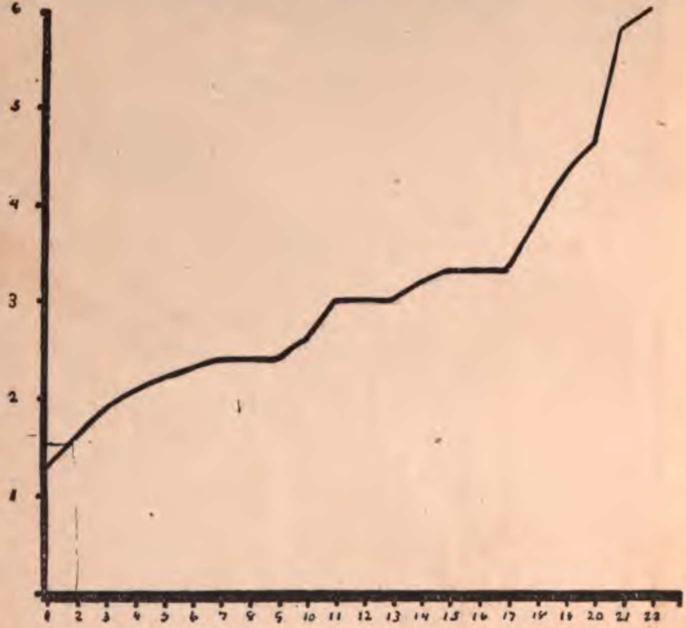


Diagram 1. Curve representing the Rate of Reading in a Third Grade of Twenty-two Pupils, School A

The scale along the base of the figure represents the numbers of the children in the grade. The scale at the left shows the rate of reading in words per second. The papers were arranged in the order of rate of reading. (H. A. Brown, 1916.)

It is possible from the data given in Tables 1, 2, 3, 4, and 5 and the *graphs* presented in Diagrams 1, 2, 3, and 4 to get an accurate picture of the condition of the class. Diagram 1 shows the *reading rate*, and it appears that there is a *variation* from 1.28 words per

---

[1] Italics in the quotations are mine.

second to 6.02. This is a larger *variation* than ought to exist in a grade, but no larger than that usually found. The *average comprehension* of the class is 40, which is high, and the *reading efficiency*, which is 114.79, is high. We find individual *variations* in comprehension and reading efficiency, but these are not nearly so great as in most of the classes thus far tested. In fact, it can be

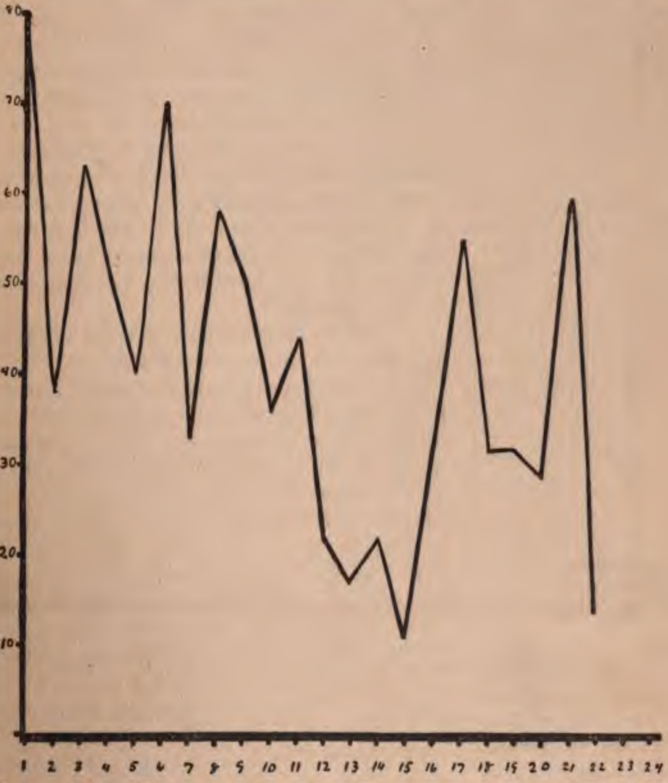


DIAGRAM 2. CURVE REPRESENTING COMPREHENSION OF THE SAME CHILDREN AS IN DIAGRAM 1 AND IN THE SAME ORDER

The scale at the left shows the comprehension. (H. A. Brown, 191–).

said that the class is in a rather satisfactory condition in this respect. While the rate of reading is high, there are, however, ten pupils whose rate is considerably below the average of the class. They should be given special quick perception practice daily to bring their rate of reading up to a higher standard. There are ten pupils whose comprehension falls considerably below the average of the class, four of whom fall conspicuously low and can easily be identified in Diagram 2. They need special practice in rapid silent reading with special emphasis upon getting a maximum of content from what is read. The four who get the lowest marks in comprehension are seen on Fig. 3 to have a very low score for reading efficiency.

We may now examine a number of individual cases. It is easy to see that Pupil No. 1 is deficient in the rate at which he can read. He gets a relatively large proportion of the content at his present rate of reading, but he reads so little in a unit of time that his efficiency is low. He should have practice to increase his speed, and if it is found that at a higher rate of reading his comprehension is poor, he should be given practice for the purpose of bringing about improvement along this line also. Pupil No. 2 has a difficulty which is easy to diagnose. In the first place his rate of reading is not sufficiently rapid, but on quantity of reproduction he stands high. His mark for quality, on the other hand, falls to zero. In other words, he gets a good many ideas in the rough but gets nothing accurately. We see in the case of this pupil one advantage of the method of scoring reading ability advocated in this bulletin. It enables us to find more correctly the exact location of defects in the reading ability of individual children. What this pupil gets is a mere smattering of the idea. His low mark for comprehension, together with his low

Test No. 1

Attempts		SCORE	Rights	
FRQ.	DEV.		FRQ.	DEV.
		24		
		23		
		22		
		21		
		20		
		19		
		18		
		17		
1		16		
0		15		
1		14		
1		13		
5		12	1	
2		11	0	
9		10	3	
7 19		9	5	
8		8	1	
5		7	3	
3		6	6	
2		5	7 19	
1/		4	4	
/45		3	7	
		2	4	
		1	3	
		0	1/45	
Ap. Med.	9.0		5.0	
Cor.	.6		.6	
Med.	9.6		5.6	
M.D.	Accuracy		= $\frac{5.6}{9.6}$	= 58%
Var.				

DIAGRAM 3. RECORDING AND COMPUTING DEVICE FOR DETERMINING CLASS EFFICIENCY IN FORMAL PROCESSES IN ARITHMETIC

Note the use of statistical methods. (S. A. Courtis, 1917.)

rate of reading, gives him a *low efficiency*. He needs to work both for *speed* and for *accuracy*. Pupil No. 4 reads at a rate considerably below the average. He gets, in a rather rough way, a very large percentage of the ideas, but he is very inaccurate.

Mr. Brown's material illustrates the use of averages, measures of variability and graphic methods for diagnosing weaknesses in school work.

2. *Scientific supervision of arithmetic*

This type of statistical device may be supplemented by some of Mr. Courtis's recording devices in the improvement of teaching in arithmetic. Diagram 3 gives a simple chart for tabulating the number of pupils attempting various numbers of problems, the number of pupils working various numbers of these correctly, the approximate median, (Ap. Med.); the correction (cor.); the true median (Med.); the mean deviation (M.D.) and the accuracy.

Diagram 4 presents Courtis's "Diagnostic Curve of Median Development in Speed and Accuracy" in arithmetic, the use of which is explained in the following quotation: —

> In Diagram 4 are drawn curves for two school systems. Curve A is for a small village school in New Hampshire. Curve B again represents the scores made by the group of 29 school districts in Boston which have been tested every year since 1912.
>
> The work in school A is very poor. Grade four falls entirely outside the diagram. Grades five and six in speed nearly equal the fourth- and fifth-grade standards, respectively, but in accuracy are way below the normal fourth-grade level. From the sixth grade on, the effect of school work is to emphasize accuracy, so that while the seventh and eighth grades approach more nearly the normal curve,[1]

---

[1] Mr. Courtis's use of "normal curve" in this connection should not be confused with the standard practice of reserving that term for the so-called "curve of error," the "normal probability curve," etc. There is great need for uniformity in practice in our statistical terminology. Such terms as "normal curve" have really become standard in our thinking and their specificness of meaning should not be clouded by multiplying terms.

the increased accuracy is obtained at the expense of speed. The eighth-grade scores are lower than those of the seventh grade, and none of the scores reach the normal fifth-grade level. Curves of this character are evidences of lack of supervision, of poor, ineffective teaching, and are far too common in country schools.

Curve B, on the other hand, indicates good quality of work and steady progress. Note that the curve lies wholly above the normal

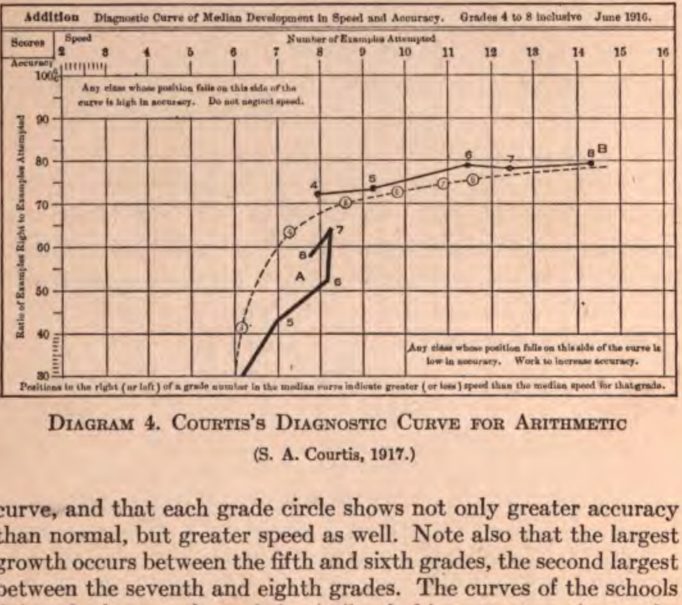


Diagram 4. Courtis's Diagnostic Curve for Arithmetic

(S. A. Courtis, 1917.)

curve, and that each grade circle shows not only greater accuracy than normal, but greater speed as well. Note also that the largest growth occurs between the fifth and sixth grades, the second largest between the seventh and eighth grades. The curves of the schools doing the best work tend, in similar fashion, to approximate the 80% line in addition, although few attain to as high speed levels as those shown in curve B.

3. *Studies of failures in the public schools*

One of the most important types of administrative study that can be made of a school system is a study of the failures of its pupils in the different grades and different subjects. Somewhat recently these problems of non-promotion have been studied analytically, to the great benefit of the schools in question. A practical graphic device for revealing lack of adjustment between pupils and the work of specific grades or subjects is shown by samples from Mr. Judd's Report in the Cleveland Survey,[1] Diagrams 5, 6, and 7.

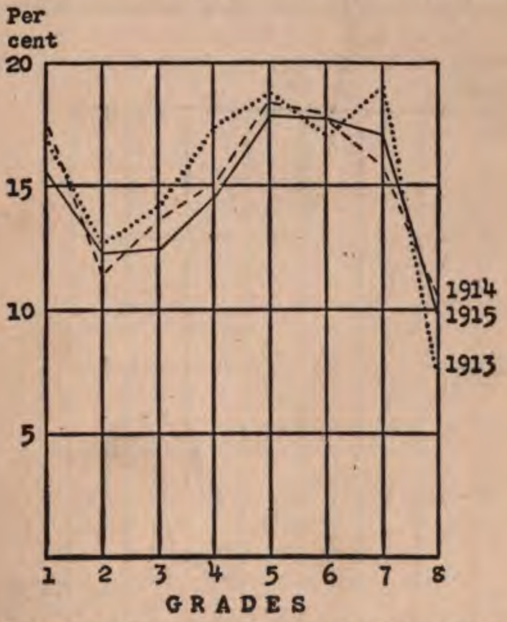


DIAGRAM 5. PER CENT OF FAILURES IN EACH GRADE FOR THREE SUCCESSIVE JUNE PROMOTIONS

(C. H. Judd, *Cleveland Survey Report*, 1915.)

4. Comparative method of analyzing city school costs

In recent years many superintendents of schools have been adopting simple quantitative and graphic methods of ac-

---

[1] Judd, C. H. *Measuring the Work of the Public Schools.* (Cleveland Survey Foundation Reports. 1916.)

quainting the public in their communities with school needs and school practice. Progressive among these has been Superintendent F. E. Spaulding, now of Cleveland. Diagram 8 illustrates his adaptation of the comparative method in studying the financial status of a school system. After a very detailed comparison of the expenditures of Minneapolis for specific school activities, with those of twenty-four other cities, he sums up the situation in the following diagram: —

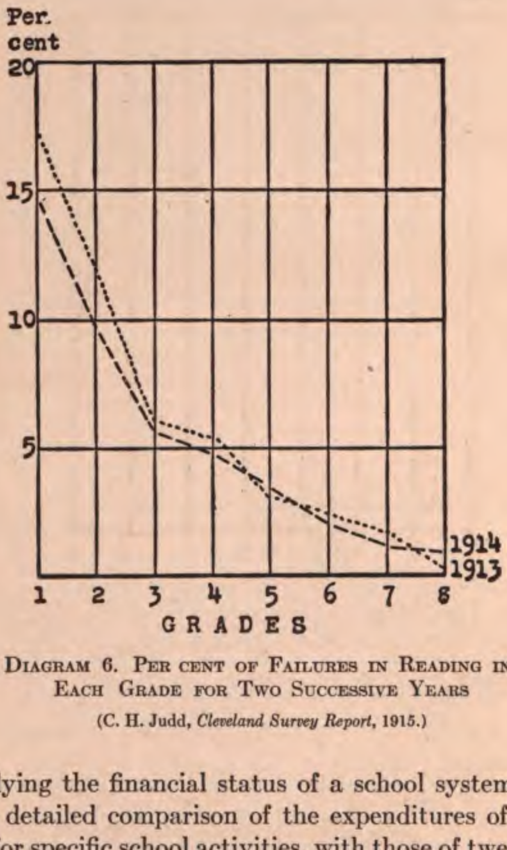


Diagram 6. Per cent of Failures in Reading in Each Grade for Two Successive Years

(C. H. Judd, *Cleveland Survey Report*, 1915.)

5. *Cost for high-school subjects*

The comparative method of studying school situations has led to the use of many statistical and graphic methods

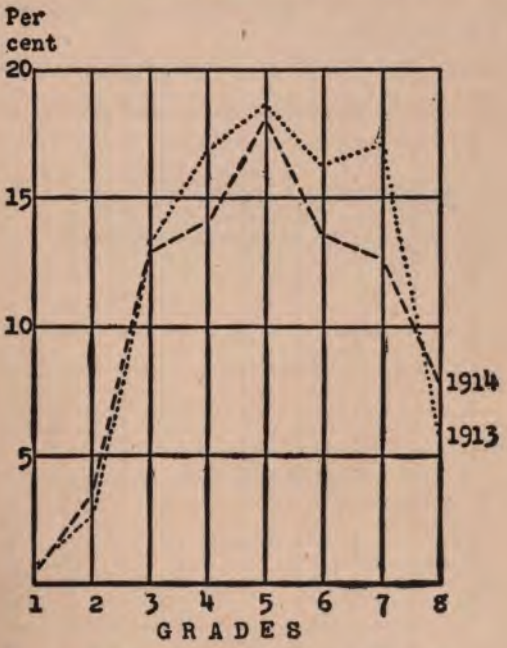


DIAGRAM 7. PER CENT OF FAILURES IN EACH GRADE FOR TWO SUCCESSIVE YEARS

(C. H. Judd, *Cleveland Survey Report*, 1915.)

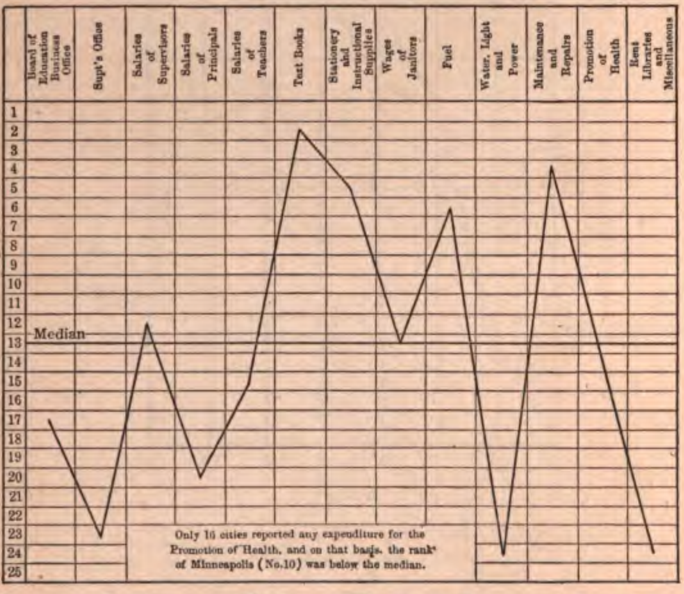
of presentation and interpretation. Mr. Babbitt's use of the middle 50 per cent (those between the two *quartiles*) to give a "zone of safety," by indicating both the attainment and relative position of each city, school, or class in the group is shown in Diagram 9.

6. *Use of "ranking" methods to determine relative standards in school efficiency*

Comparative "ranking" methods of studying school efficiency were used by Updegraff, in his *Study of Expenses of City School Systems.*[1]

In his discussion of method of treatment of data he says: —

It has come to be generally accepted that the way in which to



Only 16 cities reported any expenditure for the Promotion of Health, and on that basis, the rank of Minneapolis (No.10) was below the median.

DIAGRAM 8. SHOWING RELATIVE RANK OF MINNEAPOLIS FOR ALL SCHOOL EXPENDITURES

Is Minneapolis spending too little, comparatively, for janitors' wages? The diagram shows Minneapolis was the thirteenth or median city for this item of expenditure. Do the supervisors receive too large a proportion of the total school expenditures? Minneapolis is twelfth compared with the other cities of her class. Has the average expenditure for five years been high for textbooks? Yes. It was the second in the list. But during part of the period high schools texts sold at cost to pupils were included under general maintenance. (F. E. Spaulding, 1916.)

---

[1] Updegraff, H. Bulletin no. 5, U.S. Bureau of Education. (1912.)

give the clearest and at the same time the most accurate measure of a series of numbers is to state the *median of the series* and the

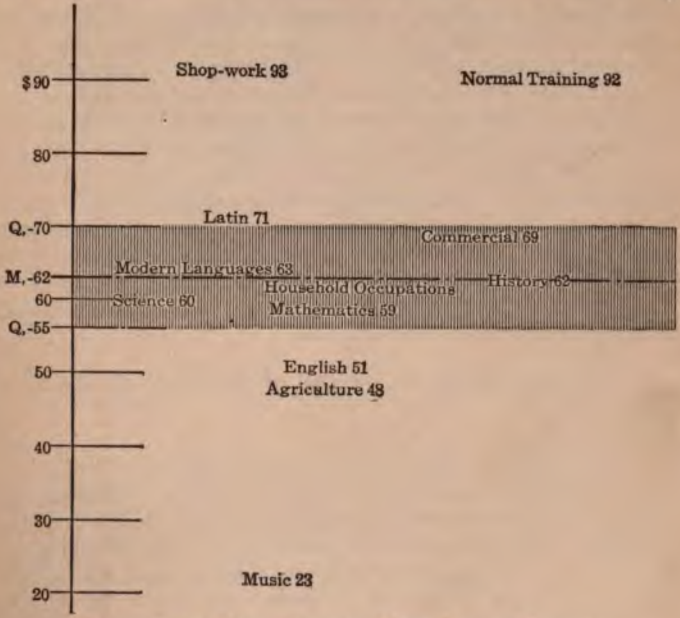


DIAGRAM 9. MEAN COSTS OF HIGH-SCHOOL SUBJECTS

"The variety of prices paid for the same quantity of instruction in the various subjects is shown. The subject of *median cost* stands at $62. The *middle zone of variability* shows a *range* from $55 to $70. For those that now stand above this zone of '*normal variability*' it is possible that administrative readjustments are desirable for the purpose of bringing them down, and thus eliminating waste. For those below this normal range of variability classes need to be cut down in size, teachers better paid, or the teaching week shortened, so as to bring them at least nearer the range of normality. In other words, just as it is possible to determine standard costs for each of the various subjects separately, out of the practical situations where those subjects are taught, so it may be possible to determine flexibility standards of cost for the entire situation applicable to the entire range of subjects. Whether or not this can be profitable can be known only after such standards have been derived for high schools of *homogeneous classes*, and involving large numbers. After the matter has been tried out its worth can be known." (F. Babbitt, 1916.)

limits of the *middle 50 per cent.* In time past the *arithmetical mean* or average has been used for this purpose, and it still has its value. Nevertheless its disadvantages, especially that of the undue weight exercised by a number which is very large or very small as compared with the others in the series, are causing the increased use of the *median* wherever practicable.

The second feature of the general method of treatment is the "*ranking*" of the various amounts in each column by groups. The "*rank*" of an item is its place in the series, as arranged for the determination of the *median* and the *middle 50 per cent*, as just described, the item lowest in value being given rank 1, the next to the lowest rank 2, and so on. In other words, the "ranks" are the result of the process of the numbering of the series, which necessarily precedes the determination of the median and the middle 50 per cent. No element of comparative worth is attached to the numbers given. In some items, as in fuel, it is creditable to a city to have a low number; in others, a high number. The purpose for the insertion of the columns entitled "rank" in the tables is merely to facilitate the comparison of items.

As an illustration of his use of the method, we may quote: —

Comparison of distribution of expenses in one city with distribution of expenses in other cities of the same group. This may be done in a cursory manner by extending the process just indicated to all items, and forming a rough judgment as to the items in which the city is low or high as compared with the group as a whole. The more accurate method consists in *computing the differences between the percentages of the various classes of expenses for the city and the corresponding medians, and arranging the excesses and deficiencies in separate lists.* As those items that vary most from the medians are of greatest importance, and as *variation* from the median to the extent of the limits of the middle 50 per cent may be regarded as normal, the computation of differences in cases wherein the city's percentage is within the limits of the middle 50 per cent may be for all practical purposes neglected. The following diagram (10), presents the result of such a computation for the city of Washington.

7. *Use of the "normal curve" in designing school tests*

In attempting to improve the marking of pupils and the planning of school tests, much recourse is had now to the *normal probability curve.* One example can be given from the writer's discussion of standardized tests in algebra.[1]

A more complete quotation from this study is given in Chapter VIII. This briefer one will serve to illustrate the method: —

DIAGRAM 10. DIFFERENCES BETWEEN THE VARIOUS PERCENTAGES OF TOTAL EXPENSES THAT LIE OUTSIDE THE LIMITS OF THE MIDDLE FIFTY PER CENT, AND THE MEDIAN PERCENTAGES FOR THE SAME ITEMS, FOR WASHINGTON, D.C.

(H. Updegraff, 1912.)

Let Diagram 11 represent the distribution of algebraic abilities in the pupils represented by our 27 school systems. The base line then represents a "scale of algebraic difficulty" ranging, let us say, from nearly 0 ability to nearly perfect or 100 per cent ability. . . . Taking as our unit of measurement on the base line, *sigma*, σ, or the "standard deviation" of the distribution (indicated graphically on Diagram 11), and laying it off 2.5 times each way from the mid-point of the curve, gives us 5 divisions (which may be conveniently divided into 10 divisions, corresponding "practically" to our public-school marking system). In doing this we are arbitrary to the extent of neglecting only 0.62 of 1 per cent of our pupils at each end of the base line. If this 0.62 of 1 per cent is thrown into the middle of the curve where the individuals are more closely grouped, it is a negligible factor. Calling the point 2.5 x sigma

---

[1] *School Review*, February and March, 1917.

from the mid-point 0, and setting the successive points 10, 20, 30, etc., to 100, we now have a practical working "scale of algebraic difficulty" over the successive points of which the corresponding percentages of our pupils may be indicated. Doing this, we see in

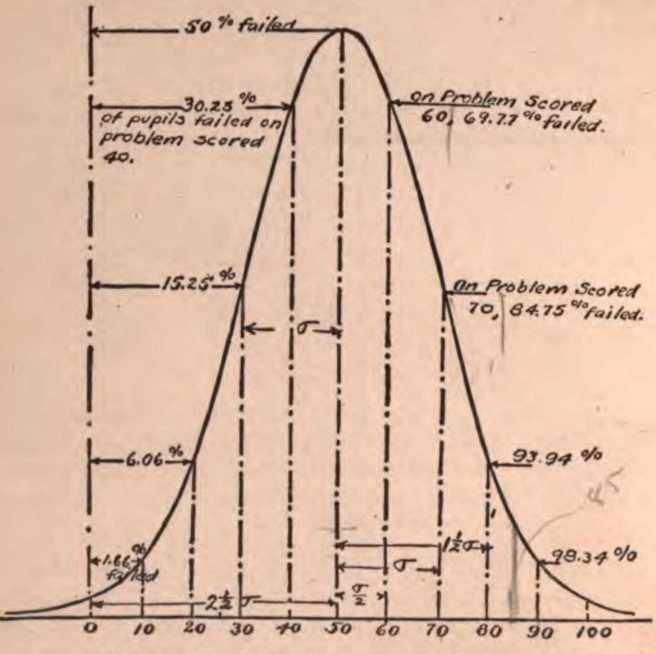


DIAGRAM 11. SCALE OF ALGEBRAIC DIFFICULTY

Distances on the base line represent, to scale, relative difficulty of problems. Area under the curve represents total number of pupils that were tested for ability to translate verbal problems. 0 and 100 points set arbitrarily at $2.5 X \sigma$ from the mean. Mean is set arbitrarily at 50. Area of the curve between 0 and any point on base line represents percentage of pupils who failed the problem placed at that point. (H. O. Rugg, 1917.)

Diagram 11 the proportions of our group of pupils that correspond to various degrees of difficulty on the base line. Thus a problem which is failed by 96.6 per cent of the group falls at the point marked 85; that failed by 84.8 per cent is scored 70, etc., throughout the list. To enable us to mark in an accurate way, a table has been computed in which the base line has been divided into 500 parts.

8. *Distribution of general intelligence in school pupils*

The study of the distribution of general intelligence in pupils in our public schools is likewise making use of quantitative methods. Terman[1] points out the symmetry of the plotted results of testing the intelligence of 905 school children, as follows: —

> The I Q's were then grouped in ranges of 10. In the middle group were thrown those from 96 to 105; the ascending groups including in order the I Q's from 106 to 115, 116 to 125, etc.; correspondingly with the descending groups. Figure 12 shows the distribution found by this grouping for the 905 children of ages 5 to 14 combined. The subjects above 14 are not included in this curve because they are left-overs and not representative of their ages.
>
> The distribution for the ages combined is seen to be remarkably symmetrical. The symmetry for the separate ages was hardly less marked, considering that only 80 to 120 children were tested at each age. In fact, the *range*, including the *middle 50 per cent* of I Q's, was found practically constant from 5 to 14 years. The tendency is for the middle 50 per cent to fall (approximately) between 93 and 108.

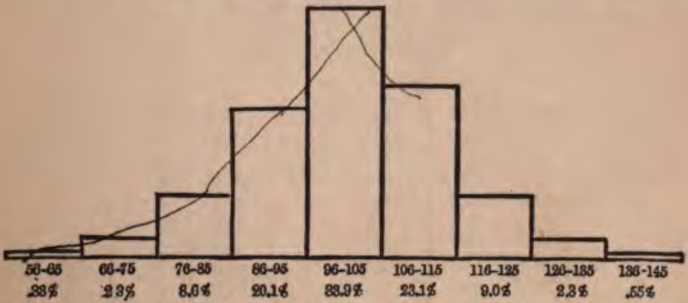


DIAGRAM 12. DISTRIBUTION OF I Q'S OF 905 UNSELECTED CHILDREN, 5-14 YEARS OF AGE

(L. M. Terman, 1916.)

---

[1] Terman, L. M. *The Measurement of Intelligence*, p. 66. (Houghton Mifflin Co., 1916.)

9. *Correlation between mental tests*

A quotation from Freeman's[1] discussion of methods of testing in the laboratory shows the following use of *correlation:* —

TABLE 2. CORRELATION BETWEEN FIRST AND SECOND OPPOSITES TESTS

Individual	*Score in I*	*Score in II*	x *diff. of scores in I from average*	y *diff. of scores in II from average*	x^2	y^2	xy
1......	15	10	−4	−3	16	9	+12
2......	15.5	10	−3.5	−3	12.25	9	+10.5
3......	16	6	−3	−7	9	49	+21
4......	17.5	10	−1.5	−3	2.25	9	+ 4.5
5......	17.5	11	−1.5	−2	2.25	4	+ 3.0
6......	17.5	18.5	−1.5	+5.5	2.25	30.25	− 8.25
7......	18.5	11	− .5	−2	.25	4	+ 1
8......	19.5	13	+ .5	0	.25	0	0
9......	20.5	10	+1.5	−3	2.25	9	− 4.5
10......	20.5	13	+1.5	0	2.25	0	0
11......	20.5	20	+1.5	+7	2.25	49	+10.5
12......	22	17.5	+3	+4.5	9	20.25	+13.5
13......	23.5	16	+4.5	+3	20.25	9	+13.5
14......	24	18	+5	+5	25	25	+25
Average.	19	13			105.5	226.5	101.75

$$r = \frac{\Sigma x \cdot y}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}} = \frac{101.75}{\sqrt{105.5 \times 226.5}} = \frac{101.75}{154.6} = 65.8$$

$$\left(\text{or } \frac{\text{sum of the products of } x \text{ and } y}{\text{square root of (the sum of } x^2 \times \text{the sum of } y^2)} \right)$$

Table 2 illustrates a form of procedure which is necessary, in many cases, to obtain a reliable calculation of *correlation*, that is, the determination first of the *reliability* of the measures secured in each test by itself. This is secured by finding the correlation

---

[1] Freeman, F. N. *Experimental Education*, pp. 177–79. (Houghton Mifflin Company. 1916.)

between the two performances in the same test, using, where the nature of the test demands it, different subject-matter in the two performances. If this correlation is not fairly high — above .60 — the degree of correlation between this test and others is of little significance, since the scores are not accurate measures of the ability in question. A formula has been developed by Spearman to correct a *coefficient of correlation* when it is reduced by lack of precision in the results in the individual tests, but the reliability of this formula is doubtful, and it is far better to perfect the methods of giving the test until their results are consistent. In the case before us two series of opposites were used with the same persons. The *correlation* between them appears from the table to be satisfactory ($r = 65.8$), though it might well be higher.

Use of quantitative methods. The foregoing quotations offer but a crude and inadequate picture of the extent to which students of education are making use of quantitative methods in attempting to solve their administrative and pedagogical problems. They merely serve to illustrate the principal statistical and graphical methods which will be taken up in the succeeding chapters. It is felt, however, that there is a need for a more complete organization of the "field of educational research" than as yet has been made. Many quantitative studies have appeared in each of the various phases of scientific education. The student is baffled by a maze of scattered material. To aid him in organizing his thinking, by cataloguing the various educational problems and the methods by which school men are trying to solve them, Plate I is included in Chapter X. On this plate the writer has attempted to give definite reference to all the studies that are of any importance to school men. The chart is so built as to indicate two important characteristics; it states: (1) who has studied each of the various problems; and (2) by what methods these persons have attempted to solve these problems. The key number attached to each name refers to the position in the complete bibliography given at the end of Chapter X. It will be noted that no attempt has been made to include the studies in the field of educational psychology. To give the student the key to this field, selected references containing complete bibliographies are given.

II. THE MORE IMPORTANT GROUPS OF SCHOOL PROBLEMS

In summarizing the discussion of this chapter let us bring in review a brief statement of each of the more important groups of school problems. To enumerate them we find: —

1. Administrative problems

Study of the curriculum. There have been concerted attempts to establish minimum essentials in the course of study in our schools, — question-blanks have been sent out covering various phases of the content of the curriculum; textbooks have been analyzed in a tabular way; judgments of specialists have been secured concerning the proper organization of subject-matter; industrial, economic, and social conditions in various types of communities have been studied with a view to adapting school practice to them.

Facts about the teaching staff. By means of question-blank methods and personal investigation of state school laws, city school-board by-laws, manuals, rules, and records, and Federal, state, and city school reports, — quantitative facts have been collected about the teacher: who she is, from what home environment she came [redacted], how much training and experience she has had; facts about her appointment, certification, salary, progress in the teaching profession, and her classroom efficiency.

Problems centering about the pupil. Personal study of individual systems, supplemented by the question-blank, has been used by private and public agencies to ascertain the status of pupils in our schools; in what way they are distributed through the elementary and secondary grades, according to relative ages; non-promotions and rates of progress through the grades; how pupils are eliminated from school; administrative devices ("promotion systems" or "plans") for adapting the machinery of the school system to the capacities, needs, and interests of the child; method of "marking" the pupils' achievement.

Status of school finance. Recently school costs and business management have been studied in this same quantitative way. Originally by question-blank, but mainly by individual investigation of school laws, charters, and records, specialists are establishing the legal basis of school finance, the status of city and state school revenues and expenditures, unit costs for education, methods of raising and apportioning school funds, and the efficiency of the business management of our city schools.

Measurement of school and teaching efficiency. During the last seven years the school world has at last turned to the construction and use of tests and "scales" to measure the results of teaching. Accompanying the attempt to study the content of the curriculum, to clarify and make definite the aims and outcomes of teaching, there has developed a most promising and important movement of educational measurement. In answer to the critics of the "older pedagogy" the newer and more scientific "educationist" is devising and using tests to measure the results of teaching in practically all of the "skill" or "formal" subjects. There are now available six handwriting "scales," of varying degrees of usefulness to classroom work; as many standardized reading tests; many discussions of measuring spelling "ability"; a fairly large and definite body of results in testing arithmetical abilities, some extensive work in the field of algebra tests, — with little or nothing done in the remaining subjects. Accompanying this material, we now have a growing body of critical data on the validity of such tests.

Furthermore, during the past five years more than fifty American school systems have been "surveyed" by groups of outside specialists — men who came into the systems in question and collected, by detailed personal investigation from the officers, teachers, and records of the system itself, sufficient facts to adequately typify the practice of education in that city. "School measurement" has seen its most thorough-going development in this school-survey movement.

Problems of central organization and administration. Even the board of education in American cities has been subjected to the same type of quantitative study. Its present status as to size, qualifications for membership, tenure, compensation, and methods of selecting board members; their functions, powers, and duties, and the way they carry on their business, have been numerically determined by both question-blank analysis and by personal study of the charters, by-laws, rules, and records of city school systems.

Miscellaneous educational activities. In the same fashion, various miscellaneous educational activities have been canvassed in a tabular way, — problems of school hygiene, medical inspection, rearrangements of the school year, etc.

All of the above types of problems are administrative in character. In each of them we have noted the recurrence of the fundamental initial methods of statistical inquiry, — the collection of educational data by either (1) question-blank; or (2) some method of personal investigation. These will be discussed definitely in Chapter II.

In addition to these outstanding administrative "problems," we must bring into our perspective of the "field of educational research" a statement of the more important experimental problems of learning and teaching. For our purposes a brief enumeration of the principal types of study will have to suffice.

2. *Pedagogical-Experimental problems*

Problems of "learning" were first studied in a controlled way under isolated conditions in the psychological laboratory three decades ago. The names of the leaders of various German schools, Ebbinghaus, Meumann, Kraepelin, Lay, etc., are linked up, literally, with scores of specific quantitative studies of isolated learning. These may be listed under the following points: —

Studies of the "practice" or "learning" curve. Data were collected and interpreted on the improvement of subjects in doing a particular mental act (*e.g.*, memorizing series of nonsense syllables); facts were collected on the *rate* of improvement, the amount of improvement, the limit of improvement, the mental qualities that conditioned improvement, changes in the rate and the permanence of improvement. Each of the studies involved the use of many quantitative methods. During the past fifteen years these studies have come out rapidly from American laboratories, and gradually have been extended to include specific types of mental work done in the class room.[1]

Formal discipline. Since James suggested the use of quantitative methods in studying the possibilities in formal discipline in 1890, thirty-odd reports have been made on the influence of training in one field of mental activity on performance in another field of mental activity. The old traditional *a priori* method of controversial discussion has given way to an experimental and statistical attempt to es-

---

[1] For fairly complete bibliographies on the "Practice Curve," "Mental Fatigue," "Mental Work," and "Mental Types," see Thorndike's *Educational Psychology*, vol. II, entitled *The Psychology of Learning*.

tablish scientifically the status of the possibility of "transference of training."[1]

Mental work and fatigue. In the same way the conditioning factors of "mental fatigue" and "mental work" have been tested under controlled experimental conditions, and a fairly large body of data collected.

General intelligence and mental inheritance. A very voluminous literature is already available giving the results of the application of experimental and statistical technique to this group of problems. Similarly, many studies have been reported on problems of mental inheritance, carrying over the same statistical methods from the field of biological investigation.[2]

These, then, are the administrative and experimental problems which the school man of to-day is trying to solve. During the past ten years he has turned decidedly to quantitative methods in studying school practice. Each phase of school work is being subjected to "counting" methods of study. School discussions are becoming thoroughly factual.

III. Steps in Educational Research

In revealing the problems of school research we have pointed out the outstanding *methods of collecting educational data.* At this point the student should have in mind at least a rough perspective of the general steps in the complete procedure of working out a statistical problem. In

---

[1] For a complete summary of all published literature see the present writer's *Experimental Determination of Mental Discipline in School Studies.* (Warwick & York, Baltimore, Md., 1916.)

[2] Complete bibliography on these fields of study can be found in Thorndike (referred to above); Whipple, G. M., *A Manual of Physical and Mental Tests* (2 volumes, Warwick & York, Baltimore, Md., 1916); and Stern, W., *Psychological Methods of Testing Intelligence* (Warwick & York, 1916); or in Meumann, E., *Psychology of Learning* (1916).

bringing to a close this introductory discussion we should now connect this first step in school research with the remaining steps. To merely enumerate them at this point, a complete statistical analysis of a carefully-defined educational problem would necessitate the following steps: —

A. Necessary steps.

1. The careful definition of the problem.
2. The collection of educational data.
3. The original tabulation or arrangement of data.
4. The systematic classification of data (in frequency distributions).
5. The summarization or condensation of data. Two general methods: 1. analytic; 2. graphic.

B. Analytic methods. These are classified as: —

1. The method of "averages" — representing the typical condition or "central tendency."
2. The method of "variability," representing the extent to which data *vary around the average.*
3. The method of relationship between various sets of data.
4. The method of reliability — establishing the amount of dependence that one may place on the statistical results of his investigation.

C. Graphic methods or the reporting of school facts.

The use of various types of frequency curves, diagrams, charts, etc.; the application of "type" frequency curves (*e.g.*, the normal probability curve) to educational data.

These steps and methods will be taken up and explained and illustrated in the chapters which follow.

CHAPTER II

THE COLLECTION OF EDUCATIONAL FACTS

If a superintendent of schools or an "interested citizen" wished to collect facts on any of the types of problems mentioned in Chapter I, he would have access to four principal sources of original data. These may be stated, in tabular form, as follows: —

I. General Sources of Original Educational Data

These general sources may be enumerated under the following main headings: —

A. State school laws and city board of education charters.

B. Published official reports.

  I. Federal reports, generally published annually.
    a. Annual reports of the United States Bureau of the Census.
    b. Annual reports of the United States Bureau of Education.
    c. Annual reports of the United States Bureau of Labor Statistics.

  II. State reports.
    a. Reports of state superintendents of public instruction (or equivalent officer), or state boards of education, in each of the States.
    b. Reports of other state departments: *e.g.*, Indiana Bureau of Statistics; state census reports; etc.

  III. Publications of city school systems.
    a. Manuals, by-laws, and rules and regulations of city boards of education.
    b. Periodic "proceedings" or "minutes" of meetings:
      1. Of city boards of education.

        2. Of permanent and special committees of boards of education. (Former are published in medium-sized and larger systems; latter are not.)
    c. Annual reports of city boards of education.
    d. Special bulletins, issued either by the superintendent or by some other school official, or, in a few cities, by the bureau in charge of school research.

C. Types of school research by private agencies that may contain "original" data.

I. School survey reports. Published reports are now available for forty to fifty cities, and eight States, few of which, however, contain "original" data. Material mostly of "summarized" and comparative type.

II. Published reports of studies made by educational foundations (*e.g.*, Russell Sage Foundation, Division of Education; Carnegie Foundation for the Advancement of Teaching; General Education Board).

III. Published reports of studies made by individuals, containing, in rare cases, "original" data.

D. The original records of:

    I. Federal and state bureaus or departments.
    II. City school systems.

These, then, are the sources[1] which are now available for the collection of facts about educational practice and conditions. It will be of some value to describe briefly each of

---

[1] Each student of school research should also secure, each year, a bulletin issued by the United States Bureau of Education, entitled *Educational Directory* (for 1915–16, 1916–17, etc.). This publication contains complete lists of the names of officers of (1) United States Bureau of Education; (2) state school systems; (3) state library commissions; (4) superintendents of schools in cities and towns of 2500 population and over; (5) associate and assistant superintendents in larger cities; (6) county superintendents; and (7) officers of miscellaneous institutions; *e.g.*, schools of pedagogy, normal schools, colleges, and universities, schools for blind and deaf, feeble-minded, etc., schools of art and of industry, parochial schools, directors of museums, library schools, church educational boards and societies; state, national, and international educational and other learned and civic organizations.

the types of data that can be secured from each source, naming the kind of facts to be found, and characterizing the relative validity of each.

A. SCHOOL LAWS AND CITY SCHOOL CHARTERS

At the present time a codification of the state school laws (a summary of all legislation affecting the conduct of schools in each State) is issued by the Department of Education (or of Public Instruction) in nearly every State. Those desiring to collect detailed facts on the legal status of any aspect of school administration should turn to these sources. Various compilations of state legal provisions, and decisions of state and federal courts on school matters, have been made under the direction of the United States Bureau of Education. Bulletin no. 47 (1915), *Digest of State Laws Relating to Public Education, in Force January 1, 1915,* is a rather extensive compilation of the actual legal basis of American school administration. In addition to this, the United States Bureau of Education has issued each year a compilation of legislative and judicial decisions on education for the current year. Of all these sources, the codifications of the state school laws themselves are the only ones containing the detailed legislation.

City school-board charters are found in various published sources, sometimes published and bound with certain issues of the annual report of the board of education; more commonly published and bound with the rules and by-laws of the board. Thus they are quite generally reprinted only on dates of revision.

Students who desire to study the legal basis of any aspect of city or state school administration should turn to the original statement of the law itself, found in one of these sources.

B. Published Official Reports

1. *Federal reports*

Educational statistics have been published annually by three federal agencies: the Bureau of the Census, the Bureau of Education, and the Bureau of Labor Statistics. Let us characterize each of these briefly.

(a) **Educational statistics in reports of the United States Bureau of the Census.** Prior to 1912 this bureau published completely analyzed data on public-school finance. The most immediate sources were found in an annual bulletin called *Financial Statistics of Cities*, and covered all American cities of 30,000 population and over. The published facts included complete descriptions of methods of securing the data, of the accounting terminology used by school statisticians, detailed statistics of receipts and disbursements, property valuations and municipal indebtedness for all city departments including schools, classified in such a way as to permit intelligent study of school costs.

These data, as reported to and including the year 1911, were collected by agents of the bureau by personal tabulation from the records of the school systems in question. Data on school cost, to be comparable, must be classified on a uniform basis. Prior to 1911 it was a very evident fact there was no semblance of uniformity in the accounting methods of different city school systems. Hutchinson in 1914 reported that he visited thirty-eight cities trying to secure comparable data on school costs, and found the summarized statistics worked out on so many different bases that it was impossible to make comparative statements about the cost of different kinds of school service and school activities from these summary statements. The agents of the Bureau of the Census, therefore, rendered a distinct service in classifying, in detailed fashion, and on pertinent educational bases, various educational financial data. The best assumption the student can make about the validity of original administrative data on school costs is that those in the reports of the Bureau of the Census are approximately accurate. The relative validity of these data and those in the reports of the United States Bureau of Education will be discussed below.

In addition to the purely educational statistics that can be found in the *Financial Statistics of Cities*, in detailed form through 1911, and in general summary form since 1911, the Bureau of the Census published many reports containing municipal, economic, population, and industrial statistics. Various special bulletins can be secured by addressing the Director of the Census, Department of Commerce and Labor, Washington, D.C.

(b) **Annual Reports of the United States Bureau of Education.** The Commissioner of Education publishes each year an annual report, in two volumes. Volume 1 contains descriptive summaries of educational movements, past and present. Volume 2 reports detailed statistics of all phases of public and higher education in this country, for cities and towns of 2500 population and over. These include all facts on school finance analyzed in very detailed fashion, facts on the distribution, grades, experience, training, age, sex, and pay of teachers; facts on attendance, enrollment of pupils in public and higher special schools, etc. In addition to these the bureau also publishes, intermittently, compilations of original statistics covering particular aspects of school administration, as, for example, salaries paid to various grades of teachers, together with number of teachers receiving these salaries; salary schedules, etc., for all cities above 2500 population, etc.

Validity of data in reports of the Bureau of Education.

These data have always been secured by question-blank methods; almost never by personal investigation of the records of the school systems by agents of the bureau. They are collected annually on a detailed blank form, the business and statistical clerks of the various systems filling in the required data. The result of the use of this method has been that the statistics have been very unreliable (for comparative purposes), both absolutely and relatively. Prior to the year 1911 they were distinctly so, due to the fact that there was almost no uniformity in city school accounting methods, and there was comparatively little agitation (at least prior to 1905) for getting cities to use uniform systems of records and reports. During the years 1905 to 1910, a growing demand for improvement in these conditions led to the coöperation of the United States Bureau of Education, the National Education Association, and the newly formed National Association of School Accounting Officers (1910) in an attempt to standardize accounting and statistical methods in city schools. A joint committee of these agencies recommended the adoption of a "Standard Form," for recording and reporting all types of school statistics. The Bureau of Education adopted this form in 1911 for its annual collection of data, and a decided improvement has taken place in the character and validity of the school statistics during the past five years. It is estimated that fully five hundred American city systems are now classifying their records in accordance with this form. It is true, however, that many cities, particularly some of our larger cities, having school officers of initiative and originality, have been slow to change their school accounting systems to accord with the standard scheme. Even to-day some of these, although laboriously retabulating their statistics for the Commissioner's Report each year, use their own independent system of accounting.

Thus, it is believed that, since 1911, the educational statistics of the Bureau of Education have increased steadily in reliability for "comparative ranking purposes," although still collected by question-blank methods. It is to be regretted that, with the use of the standard form by the Bureau of Education, the Bureau of the Census stopped making its detailed classification of educational statistics in 1911, reporting since that time only very general summaries of school receipts, expenditures, indebtedness, etc.

In making the study of the *Public School Costs and Business Management in St. Louis* (1916), the writer attempted to establish the validity of the statistics of the Bureau of Education for purposes of comparing various cities by arranging them in "rank" or "serial" order in their various financing activities. It was assumed that the financial statistics of the Bureau of the Census to and including the year 1911 were approximately correct. It was found that the Bureau of Education in the same year, 1911, published the same type of financial statistics, thus providing an opportunity for direct comparison of the absolute figures compiled by two agencies on identical school activities. Tables in the complete survey report give, as obtained from each source, the total expenditures and differences in amount spent for each of a list of cities, for nine different kinds of school service — special supervision, principalships, instruction, supplies, etc. Tables computed and stated in the survey report give the *per pupil cost* for each of these nine kinds of service, together with the rank of each of the cities in the group for each item. It is clear from inspection of those tables that we have to discuss the validity of the data as collected from these two sources strictly in terms of the use we are going to make of them. *First:* if we merely are going to rank cities in terms of per pupil cost, then the statements made in the survey report are valid, namely: —

With few exceptions the tables show a very satisfactory agreement in position, the costs for supervision and principalships being the ones for which less agreement would be expected than for any other activities. The conclusions that we form from one set of records will not be unlike those formed from the other set of records. Especially is this true in the case of the one city in which we are interested, St. Louis. We may summarize its position in all the tables as follows: —

	Salaries of					*Textbooks*
	Supervisors	*Principals*	*Teachers*	*Repairs*	*Janitors*	
Bureau of Census..........	4	7	11	8	4	8
Bureau of Education.......	4	6	9	6	3	8

The largest displacement in the ranking for St. Louis is two places. As a result of the tabulation and ranking it is believed that the interpretations made on the financial situation in St. Louis, from cost tables computed from the Annual Report of the United States Bureau of Education, 1915, will be valid. Especially is this true since 1912 was the first year in which the bureau collected statistics on the standard form, and much improvement has come about since in the completeness and accuracy with which city systems report their school facts.

The most frequent use that school men want to make of educational statistics is of this very "comparative" and "ranking" type. One point should be noted, however. These cities are the largest cities in the country, and have the most thoroughly equipped accounting and statistical staffs, supervised by specialists in this field. The experience and investigations of the writer lead to the belief: —

(1) That considerable reliance may be placed on the comparability of the classification of educational statistics for groups of medium-sized cities (15,000 to 40,000 for example). These cities are following the "Standard Form," even more

closely than are the larger cities. The comparative financial statistics of the Bureau of Education for twenty-one cities in Indiana, Illinois and Wisconsin (between the sizes of 15,000 and 25,000, and within 150 miles of Chicago) have been checked with care. The results show a fair agreement between the records as compiled by the bureau and by other agencies. The methods have been checked personally in three of these cities, and show that considerable reliance may be placed on the absolute expenditures reported, as well as on the "position" of each city in the group.

(2) In the study of larger cities, however, it was found that, if we wish to deal with the absolute statistics of cost, attendance, teaching staff, etc., we must make decided mental reservations in our acceptance of the Bureau of Education figures. In the first place, there are occasionally very large differences in reported figures due to incorrect classification (for example, expenditures for supervisors and principals in certain cities). In the second place, differences of 10 to 20 per cent are relatively common in these tables. The present study, however, can merely warn the student of the large inaccuracies in the absolute figures reported by certain cities to the Bureau of Education.

(c) **Annual reports of the United States Bureau of Labor Statistics.** If the school man desires statistics on the occupational situation, distribution of workers as to grade, trade, salaries paid, etc., he can find such data in annual reports and bulletins of this bureau, by addressing the director.

2. State school reports

The superintendent of public instruction, or the department of education in each of the states, now issues either biennial or annual reports on educational activities in the state. A very considerable body of original statistical material may be found in these. It is fairly common, for example, to classify the statistics by counties, instead of enumerating them for cities and towns. On the whole, it is rarely that one can find detailed data on city schools in state school reports. Furthermore, it is uncommon to find data detailed enough on town and rural schools to permit of comparative studies of educational practice in specific communities. The reports are filled up with narrative reports from county and other school officers, from various special and higher institutions controlled by the state, state courses of study, reports on county institutes, digests of school laws and legal decisions, state examination questions, and other types of descriptive material. They all give certain detailed financial and attendance statistics on the "common schools," arranged by counties. In exceptional cases, good comparative data can be obtained. For example, the state report for Missouri contains a detailed financial analysis for several hundred towns and cities in the state. It is possible to use the data in making a comparative cost study for particular communities, grouped in various ways.

3. Publication of city school systems

(a) Manuals, by-laws, rules, and regulations. All of our larger cities, and many of the smaller ones, print annually handbooks or "manuals" giving miscellaneous data concerning the administration of the city schools. They may include certain fiscal data for various city departments, and sometimes the "charter" under which the board operates; the "by-laws" enacted by the board to govern its conduct and to create a complete working organization for the schools of the city, to endow and state specifically for each of the officers his powers and duties, and to state the "rules" governing the schools. They also contain, very probably, the districting of the city system, rules governing: (1) pupils; (2) grading, salary schedules, eligibility, appointment, promotion, etc., of teachers; (3) operation of departments outside the educational department.

(b) **"Proceedings" or "minutes" of meetings of city boards and their committees.** These are now very generally printed for the larger cities, monthly, semi-monthly, or weekly. They very often are found to duplicate the fiscal facts printed annually in the school report; they often contain the detailed itemization of school facts that properly ought only to be typewritten and filed in the boards' offices (*e.g.*, financial itemization of all vouchers paid, regardless of amounts; lists of names of pupils graduating from various schools, etc.)

(c) **Annual school reports.** It is a fairly common practice now for cities of 30,000 to 50,000 population, and above, to print an annual school report. During the past ten years distinct changes have come about in the types of original data that these contain. The tendency toward standardization, uniformity, and a clearer classification of school facts is evidenced by the better organization of data. To a student desirous of making a comparative study of school conditions (say of cost, elimination and retardation, non-promotion, teaching staff, or what-not), the statement should be made that even now, with all the improvements which have been made in recent years, comparable statistics on these or other phases of school practice are not to be obtained from the annual school reports of our cities. This is true even for the very largest cities, with their well-organized accounting staffs.

The above, in the main, comprise the larger sources in which students of educational administrative problems may find original data. In rare cases one can discover original detailed statistics in studies made by individual students, either working as officers of a city bureau of school research, or in some educational institution or "foundation."

In summing up this brief discussion of the sources and validity of original data, the writer would urge the direct collection of statistics and data from the records and persons in the school systems in question. Question-blanks sent out by individuals rarely have resulted in sound comparative conclusions that benefit school practice. The tendency at the present time is for question-blanks to receive a decreasing amount of respectful attention from a much overburdened school world. When economically possible, personal collection gives much more valid results. It leads to: (1) a more consistently uniform original record; (2) a complete original record (*i.e.*, no data are suppressed); (3) thoroughly comparable bases of interpretation; (4) a more consistent interpretation of the facts as expressed in original and summary tables. Studies which demand recourse to state school laws, charters, rules, and other official state and city documents rest, of course, upon a perfectly valid basis.

II. Methods of collecting Educational Data

The source and validity of the various types of educational statistics having been discussed, we now turn to the methods by which data are collected. The analysis of these methods, as given in Chapter I, made many references to the two most important methods: (1) use of the question-blank; (2) personal investigation. We shall next take up the detailed discussion of these two general methods, turning to the question-blank method first.

A. The question-blank method of collecting educational data

Plate I shows the very great use that school men have made of the question-blank in studying their problems. There is hardly a phase of school administration that has not been subjected to that type of analysis. Present practice and conditions as to the content of the course of study have been established in arithmetic by Jessup and Coffman, and by Van Houten; in algebra by Denny and Mensenkamp; in spelling by Pryor; in handwriting by Freeman; in the high-school subjects by Koos, etc.[1] The present status of the teaching staff is tabulated from the "Standard Form" replies each year by the United States Bureau of Education. It has been studied by the use of the same method by Coffman, Thorndike, Coffman and Jessup, Ruediger, Manny, and Boice; by committees of the National Education Association and other organizations. The question-blank method has given facts on the age-grade distribution of pupils, — collected by Thorndike, by Strayer (both through the agency and authority of the United States Bureau of Education), and by Ayres, working through the Russell Sage Foundation. Data on promotion plans have been collected at various times by the United States Bureau of Education. The study of current practices in school finance by the question-blank method, by Strayer, and by Elliott, although not leading to basic comparable results themselves, has stimulated the standardization of school financial methods very much. In the same way the status of certain phases of central administration has been determined by the work, for example, of Shapleigh, working for the Public Educational Association of Buffalo, in such studies as his determination of the effect of commission form

---

[1] For details of specific references in this chapter see bibliography at end of Chapter X.

of government on city school administration — forty-eight cities — and the present status of janitorial service in city school systems.

Enough has been said here, therefore, to indicate the frequent use that has been made of the question-blank in school research. As indicated above, use of this method by persons working in no official capacity, or by an organization of the government, has done little more than stimulate discussion of present practice and the need for greater standardization. Even the Federal Bureau of Education has had no real "extractive power" in its search for school facts, and we have already indicated the large inaccuracies in its original records. However, under various conditions we shall be forced to make some use of the "questionary" in our attempt to determine the status of current practice. For that reason it will be pertinent to give here a discussion of its design and use.

1. The design of the question-blank

Principal types of question-blanks. Question-blanks for the collection of educational data can be distinguished into three classes, in terms of the source and reliability of the facts for which they ask: (1) those asking for facts in the personal information of the reporter; (2) those asking for facts to be found in school records; and (3) those asking for introspective or retrospective analysis, judgments of specialists, etc.

1. Question-blanks asking for facts in the personal information of the reporter. In the statistical studies in education many examples may be found of this type. They include questions relating, for example, to the age and sex of the teachers, number of years of training in particular types of institutions, number of years of experience in various grades of public-school work, salary received during various stages of the teacher's career, certificates held, etc. Such questions all relate to the personal history of the person reporting the facts. It is probable that more reliance may be placed on such types of fact than on any other collected by the question-blank method. They do not involve the labor, on the part of the reporter, of going to the records of class, school, or system to get the data, with the consequent chance for error in transcription and of decrease in number of returns caused by the inability of the reporter to take the time necessary to make the search.

A second sort of data obtained from the personal information of the reporter pertains to facts concerning particular phases of school practice. For example, a question-blank sent to teachers of English in high schools contained questions such as the following: —

1. Do you have a special teacher in oral composition? Yes ......; No.......
2. Do you use a text in oral composition? Yes......; No..... If so, what?......
3. Have you a printed course of study in oral composition? Yes......; No.......
4. Do you have a course in public speaking? Yes......; No ....
5. Is the work in oral composition given in connection with public speaking? Yes......; No.......
6. Do oral lessons precede work in written composition? Yes ......; No.......
7. Who selects the topics in oral composition? The student ......; teacher.......

Inquiries conducted for the purposes of getting facts concerning the content of a particular course of study, names of textbooks used, methods of grading pupils, etc., all fall within this class. Providing questions have been clearly asked and cannot be misinterpreted, data of this type should be very reliable. Question-blanks demanding information in the immediate possession of the reporter ought to result in a very large percentage of returns to the investigator. If the blank is clearly written, well planned, short, and definite, it should result in a return of two thirds to three fourths of the blanks sent out.

2. **Question-blanks asking for facts to be found in school records.** In this group we include the collection of facts, concerning, for example, the age-and-grade distribution of pupils in schools; various inquiries of specialists which demand detailed copying of records (*e.g.*, on the problem of retardation and elimination), total expenditures for various types of school activities, administration, instruction, operation, maintenance, etc.; distribution of teachers' time to various subjects; statements from payrolls, class enrollment records; total and unit costs, etc. With this type of inquiry nothing but intimate acquaintance with the aims, and full recognition of the importance of the investigation, will cause the reporter to take the time to give comparable and complete data which will lead to the improvement of school practice.

3. **Question-blanks asking for introspective and retrospective analyses, judgments of specialists, etc.** In this group are found various types of psychological question-blanks; *e.g.*, those from inquiries aimed at determining the status of methods of study. For example, a recent inquiry of this sort quotes liberally from an article on *How My Brain Functions*, and asks the reporter to check his own mental processes against those of the author, and tell him the result. The following excerpt illustrates this type: —

> Question: Often he "thinks of nothing." In this state he experiences *euphoria*, a feeling similar to that of the convalescing patient, who prefers to lie absolutely quiet. He experiences together with this intellectual lethargy a physical inertia. While in this condition he per-

suades himself "to postpone until to-morrow what he should do to-day."

Answer: Do you note a similar phenomenon in your own experience? Please state wherein your experience differs from that of Beaunis.

The early stages of the child-study movement were quite given over to the "questionnaire" method, masses of judgments being accumulated concerning child life, mental and moral activity, and growth. Such studies involved most extreme types of "judgment" questions, and as such are the farthest removed from a purely factual basis.

It is no doubt true that the compilation of data concerning particular phases of school practice by the question-blank method will be necessary for some time to come. Since governmental agencies, such as the United States Bureau of Education and the various state departments of education, have no "real extractive power" as yet, it is clear that individuals must do the work. It is also clear, as will be shown later, that few question-blank inquiries have resulted in establishing beyond a doubt the status of the particular question they were designed to study. This is largely due to the fact of hasty and incomplete planning of inquiry blanks, and lack of recognition of the many issues and difficulties arising in the carrying on of the problem. For this reason it appears worth while to discuss, somewhat in detail, the necessary steps in the carrying on of an inquiry of this sort.

2. Essential steps in school research by the question-blank method

First step: acquaint yourself with the literature covering the field of your problem. Your first duty is to know what others have contributed to the solution of such problems. Read carefully, take notes, and make many comments on every study made in your field. Many studies have contributed little because of this very lack of acquaintance with what other workers have done. In this way needless duplication will be avoided, needed repetition will be secured, and the mistakes and the excellencies of others' research will be utilized to advantage. Our great need is to have the vital gaps in our knowledge filled in. The careful study of the literature of a specific field of work will lead to the selection of the exact problem upon which research is most urgently needed.

Second step: specific definition of the problem. The success of your investigation depends upon the clearness with which you recognize the exact problem at hand, — especially its educational implications. Write out a very specific and detailed statement of it. Visualize the carrying on of the study from the first step to the last. Ask yourself at every turn — what has this to do with school practice? What kind of facts shall be collected to throw light on this point? Does *this* point really belong in this inquiry? Plan in a rough way the tables to be made up as a result of sending out the question-blanks. In a word — project yourself through the entire investigation in order to be able to start with a perfectly clear idea of what you are to study. It is probable that a most specific definition of your problem can come only after you have read the literature on the subject, and after you have actually worked through, at least in a preliminary way, the design of the question-blank itself.

Third step: exact delimitation of the extent of the inquiry. Your study of the literature and your attempts to define your problem should lead to an exact determination of the points to be covered and the questions to be asked in your study. Plan the number and kinds of questions to be asked in the light of a careful estimate of the labor of tabulation and summary of results. Decide the number of replies needed to establish definitely the status of your problem. In doing this, count on a return of from one third to three fourths of the blanks sent out, depending on the length of the blank, the possibility and ease of giving the information on the part of the reporter, and the clearness with which the pertinency of the investigation to the needs of school practice is recognized by those to whom blanks are sent. In deciding the number of blanks to be collected, make use of methods of determining the minimum number of cases, such as are described in Chapter VIII.

Secure enough cases to satisfy statistical "criteria of reliability," and no more than are necessary to secure acceptance of the results of your inquiry by the persons to whom you will present them. One study known to the writer involved the collection of 30,000 blanks, the original tabulation alone of the returns from which would have taken at least 700 hours of clerical labor. Careful study showed that the same conclusions could be derived from the tabulation of one fourth as many cases, with the reliability of the investigation established at every point. Furthermore, the delimitation of the extent of the study calls for careful weighing of the relative value of having a small number of questions and a large number of replies, or of having a large number of questions with a small number of replies.

Fourth step: design of the questions on the blank. Nothing is more important to the success of the study than the careful placing and wording of the questions. The most detailed analysis should be made of each one. Ask yourself concerning each one: Is this question so worded that the reporter *cannot* misinterpret it? Has every term been clearly defined, so that the returns from different reporters will be exactly comparable? Is the question ambiguous? Can it be answered by Yes, No, a phrase, a number, or a check mark? Has the person who will answer this question the information desired? Is there sufficient space allowed for the most complete answer desired? Will the questions lead to specific quantitative statements? Are they factual? Have I eliminated all confusion that might arise because "factual" and "judgment" questions have been put together? A fundamental point to be kept in mind in this connection is: Can the replies to this question be tabulated so that the data can be definitely summarized and interpreted? Still better, can the data called for at this point be more completely secured by tabulation on the question sheet itself? Such points will be illustrated thoroughly in the next section.

Fifth step: design of the originai tabulation forms. Chapter III will take up in detail the tabulation of educational data. It should be pointed out here that an absolutely essential step in the design of a sound question-blank is the preparation of the forms upon which the original tabulation of the data is to be made. This means the planning of the specific headings of the tables to be compiled, and will require definite decisions concerning the arrangement of questions and the probable types of returns to be secured. Preparation of the tables will lead to a clear-cut, logical arrangement of questions, so put together as to facilitate a clear presentation and discussion in the report. A little time spent at this stage of the work will aid much in the later organization of the completed discussion.

Sixth step: preliminary collection of data on tentative question-blanks. Having decided on the wording and arrangement of the questions, collect some data for preliminary analysis of your blank and tabulation forms. Have your blank mimeographed, making say 20 to 30 copies, and ask members of the group to whom will be sent the final inquiry to fill in the questions. Tabulate these returns on your forms, and note the difficulties of tabulation and errors in interpretation of the questions. Only in this way can your blank or your forms be made thoroughly usable. Careful study of the returns will enable you to revise both the wording and the arrangement of the blank. Mimeograph it again and try it on another group, tabulating the returns. Revise once more and prepare the final copy for the printer. In selecting a group to fill in the preliminary blank, take the persons entirely at random (*e.g.*, arrange them alphabetically and take every *nth* one). This will enable you to foretell from the returns, roughly, the proportion of the entire number of cases that you can expect to receive and will aid you in deciding how many blanks to send out.

Seventh step: preparation of printed blank. If the investigation is at all extensive the blank should be printed. Practical criteria of handling and filing returns should control the selection of the material to be used. If financially expedient and practically possible, use a light-weight card instead of paper. If this is done use standard sizes, either 3 by 5 inches, 5 by 8 inches, or 8½ by 11 inches. This will facilitate filing the returns later.

These, then, are the necessary steps in the design of a sound question-blank: — know the literature concerning the problem; define the problem specifically; limit the extent of the inquiry carefully; scrutinize minutely each question included on the blank; design the forms upon which the original tabulation is to be made; organize the tentative question-blank, and try it on 20 to 30 persons; tabulate the returns, revise the blank, try it on another group, and tabulate again; print the final copy on standard-sized material, using cards where possible.

3. Guiding principles concerning the content and form of the question-blank

There are many important points concerning the selection of questions and the form of the blank that need to be commented upon before leaving this question.

1. **" Factual " questions.** A fundamental principle for the selection of questions is that they must be as "factual" as possible: Thus, questions should involve a minimum of "judgment," discrimination, or "deferred memory" on the part of the reporter. For example, in this question, asked in an inquiry on the economic condition of the members of the general teaching staff of the country: —

> Check the item that would most nearly represent the parental annual income when you began teaching: —
> $250 or less,
> $250 to $500, etc.

The answer demanded memory of a situation many years past, in addition to the calculation of various items entering into the answer. The data obtained must be of very questionable value.

2. **Difficulties with " general " questions.** Education question-blanks have abounded in "general" questions. One type is the sort that nearly always can be answered "Yes," while at the same time it is nearly impossible to reply more specifically. For example, in a recent state survey of commercial education we find such questions as: —

> Do you have difficulty in obtaining clerical help?
> Do you find pupils, 14 to 18 years of age, who come from elementary and high schools, deficient in general education?

Another type of the "general" question is that which leads to unanalyzed and practically unanalyzable statements. It tends to hide up the specific facts out of which it might be possible to construct a valid general statement. To illustrate, we quote a question on the cost of teacher's education, asked in a recent survey: —

> Estimated cost of your education beyond the high school, including specific items, as cost of tuition, books, board and room, etc. . . . . . . ; and estimated cost of time as measured by the amount that you could have earned at productive employment during this period of training; . . . . Total . . . .

The answers to "general" questions seldom can be tabulated and definitely interpreted. It is a safe rule to follow that data which do not lend themselves to tabulation and statistical treatment are of negligible value to the investigator. That is, answers should be definite and susceptible of tabular classification and this should be a controlling criterion in the planning of questions. Such questions as those below, taken from a "study" of the course of study, hardly render themselves subject to that kind of treatment.

1. In what way (if at all) is your teaching of the following subjects determined by the peculiar needs and opportunities of the local community or district served by the school: —
   Agriculture. . . . .
   Manual training . . . .
   Arithmetic . . . .
   Geography . . . .
   etc.
2. What in general is the attitude of the parents toward "home work" in school studies. . . . . . .
3. What is the attitude of your community toward: —
   (*a*) Taking pupils on excursions to study neighboring industries, etc. . . . . . .

Many of these "general" questions demand of the reporter a type of discriminative judgment that but few people possess, and those only specialists trained to that particular thing. To illustrate: —

What difference in training do you notice between public high-school commercial graduates and graduates of the common private business colleges? ......

3. **Ambiguity of statement.** Many difficulties in tabulation and interpretation arise from ambiguity of statement of the question. For example, the following question, on which thousands of replies had been collected from elementary public-school teachers, had to be eliminated from the study because of the confusion in interpretation of the word "school" by many of those reporting.

Total number of pupils in the entire school ......
Number of teachers in the school, including superintendent or principal if he teaches ......
Number of pupils in the high-school department ......
In the grades ......

The returns indicated that a large proportion of the reporters had interpreted "school" to mean "school system." Many teachers from the same system reported on the same conditions, thus permitting a check. Of course, the question should never have been asked of teachers at all, but of the administrative officer.

4. **Information difficult to obtain.** Apropos of asking for facts in the immediate personal information of the reporter, it will be recognized that we must not ask for general facts that the reporter cannot give without considerable search on his part. For example, we find on a question-blank concerning the distribution of workers in certain occupations, sent to the superintendent of schools of the city in question, the following: —

1. Number of children in your community between 14 and 16 years of age at work or idle. ......
   What are they doing if at work ......
   Number of families of this whole out-of-school group to whom this income of the youth is necessary. ......

> 2. State the number of workers (in this pursuit), male and female, with different ages; the number 14 years old, ..., 15,......, 16 ......, etc., up to 80. The number of years of schooling of each of these workers by age and sex ......, etc.

The impossibility of the reporter in question filling in these data is evident.

In this connection it is clear that we should not ask for data which cannot be given accurately and in detail by the reporter, when at the same time the detailed information is available in printed records. For example, this question, on a blank directed to each of the teachers in various systems, should not have been asked: —

> State the population of the village, town, city, or district in which you teach.......

5. Other types of information. If you desire to compute *percentages*, plan the questions concerning the *number of items* so that percentages can be worked from the collected data. For example, desiring to know the *proportion* of brothers and sisters who lived to adulthood, one investigator asked: —

> How many brothers and sisters lived to early adulthood or longer?

He omitted to ask for the *total* number of brothers and sisters.

In studying problems involving many stages of growth or progress one must be careful to include *all* the possible stages or grades. A detailed question of this type is: —

> You attended country district school ...... years, village school ...... years, city graded school ...... years, one-teacher high school ...... years, larger high school ...... years, private academy ...... years, normal school ...... years, military school ...... years, college or university ...... years, graduate school ...... years, etc.

4. Rules governing the form of question-blanks

In concluding the discussion on the design of question-blanks there are certain rules of form that well may be set down: —

1. State the questions specifically. Beware of general headings or word or phrase captions. Use complete sentences or phrases long enough to convey your exact meaning to your most careless reporter. Discount the ability of your reporter to discriminate and interpret what is meant. Define and, if necessary, redefine each term which is in any respect technical, or which possibly can be misinterpreted by the least intelligent reporter in your group.

2. Plan the introductory or explanatory paragraph so clearly and completely that it will acquaint the reporter fully with what you are doing and enlist his interest in your problem. Be careful to show the pertinency of your inquiry to the improvement of *his* conditions or at least to the improvement of school practice in some particular. If you cannot do this your investigation is of doubtful value, and coöperation will not be given you.

3. Questions of arranging the *form* of the sheet are very important. Striking defects of nearly all question-blanks are (*a*) lack of clear organization of questions; (*b*) lack of sufficient space for answers; (*c*) lack of tabular schemes by which the reporter can give numerical data.

4. If the tabulation forms are designed in advance, and the complete plan of the report is sketched, the questions will be systematically organized on the sheet with this view. Arrange them in the order in which you wish to tabulate and to discuss the points of your report. Logical organization at such early stages of the procedure will enhance the clarity of your later discussion.

5. Plan sufficient space for the longest possible answer to

the question. This can be done effectively by insisting on a preliminary filling-in of the blanks. If you do this you will be almost sure to redesign the blanks in order to give longer spaces. It is rare that question-blanks are well planned in this particular.

6. When asking for continuous numerical data, provide a tabulation form on the question-blank upon which the reporter can fill in the data. Plan this tabulation form very carefully, so as to prevent errors in interpretation and in subsequent tabulation. To illustrate this point, a portion of a question-blank on junior high-school costs is quoted herewith: —

> V. Omitting all names, will you give the individual yearly salaries that were paid junior high-school principals, teachers, supervisors of special subjects, and principals' clerks during the school year 1914–15.
>
> To make it easier for you, the salaries are arranged in groups in one column. In the opposite column (marked "No. receiving"), will you place the number who received the salary stated?
>
> Example. If four women teachers and one man teacher receive *annual* salaries between $800 and $825 respectively. enter them thus: —

Annual salaries	*Men*	*Women*
800–825	1	4

Principals' salaries	*Number receiving each salary given*		*Teachers' salaries*	*Number receiving each salary given*	
	Men	*Women*		*Men*	*Women*
1000–1099			500– 549		
1100–1199			550– 599		
1200–1299			600– 649		
1300–1399			650– 699		
1400–1499			700– 749		
1500–1599			750– 799		
1600–1699			800– 849		
1700–1799			850– 899		
1800–1899			900– 949		
1900–1999			950– 999		
2000–2099			1000–1049		
2100–2199			1050–1099		
2200–2299			1100–1149		
2300–2399			1150–1199		
2400–2499			1200–1249		
2500–2599			1250–1299		
2600–2699			1300–1349		
2700–2799			1350–1399		
2800–2899			1400–1449		
2900–2909			1450–1500		
3000–3500					

Principals' clerks or other administrative clerks	*Number receiving each salary given*		*Supervisors of special subjects*[1]	*Number receiving each salary given*	
	Men	*Women*		*Men*	*Women*
250–299			700– 799		
300–349			800– 899		
350–399			900– 999		
400–449			1000–1099		
450–499			1100–1199		
500–549			1200–1299		
550–599			1300–1399		
600–649			1400–1499		
650–699			1500–1599		
700–750			1600–1700		

---

[1] For example, supervisor of art, music, etc.

B. METHODS OF PERSONAL INVESTIGATION

The foregoing pages have set before us essential principles and methods to govern our practice in the collection of educational facts by the use of question-blanks. It is undoubted that the more important contributions to the improvement of school practice will come through personal collection of facts and actual contact with the school situation itself. We should next bring in review ways and means of utilizing such methods. Since these make such complete use of tabular analysis we will take them up in connection with the tabulation of educational data, in the next chapter.

CHAPTER III

THE TABULATION OF EDUCATIONAL DATA

As students of education have turned to quantitative methods of solving their problems, the use of "questionary" methods of collecting facts has rapidly given way to intensive personal investigation. Plate I shows semi-graphically the extent to which such methods have been used to establish the status of various types of problems. A brief summary of them may be given at this point.

I. METHODS OF PERSONAL INVESTIGATION OF EDUCATIONAL PROBLEMS

A. Statistical compilation from printed material. Under this heading we have: —

1. *Tabular analysis of provisions of public school laws and city charters,* to establish the legal status of various administrative problems. For example: current practices in the various states concerning the certification of teachers; constitutional and state-school-code provisions for the administration of education in rural and city districts; the composition, methods of selecting, powers, tenure of office, and compensation of boards of education; methods of raising and apportioning school funds.

2. *Tabular analysis of rules, by-laws, and manuals of city boards of education.* By this method we determine the status of the appointment, pay, and tenure of teachers and other employees; the powers and duties of committees of the board and of its officers, and the basis of carrying on the instructional and business activities of the system.

3. Tabular analysis of textbooks and printed courses of study, to determine the present status of the content of the course of study and the relative efficiency of the order of presentation of topics in various subjects of study.

4. Tabular analysis of the data given in federal, state, and city official reports. The types of data, validity of each, and problems which can be treated from these have been taken up in the previous chapters.

5. Tabular analysis of the data found in the records of city school systems. Only by personal tabulation of facts from such records can we expect to make real progress in making known the facts on present school practice in this country. The school "survey movement" in this connection is lending great impetus to the work. Detailed comparative analysis of groups of cities is establishing definitely: — facts on the teaching staff; age-grade, elimination, and retardation facts on the pupil; facts on the marking system; detailed and systematic compilation of facts on revenue, expenditures, and unit costs; and facts on the central administration and business management of public schools.

6. Tabular analysis of facts in experimental and statistical descriptive literature.

B. Tabular analysis of results of experimentation. Under this heading we have: —

7. Tabular analysis of the "abilities" or "achievements" of pupils, through the design of "mental" or "educational" tests.

8. Tabular analysis of facts from experiments in "learning," "mental discipline," etc.

9. Tabular analysis of facts concerning the efficiency of teaching secured through the personal observation of teaching, with or without the aid of "efficiency score-cards," or schedules of "qualities of merit in teaching."

Systematic tabulation of facts. It will be noted that the collection of data in the "scientific" study of education is either: (1) straight statistical compilation of facts, from various principal sources; or (2) dependent upon the preliminary setting up of auxiliary devices for measurement (standard tests, score cards, etc.), and the conducting of experimentation. Either procedure necessitates the same fundamental auxiliary method: the systematic tabulation of facts. Experience has shown that the thoroughness and insight displayed in planning and carrying through the original tabulation is an important factor in determining the success of the investigation. We have seen already the necessity for planning the scheme of tabulation in detail at the time of designing the question-blank. The two steps in the general research thus must be carried on together — the efficiency with which one is done contributing to the success of the other. Although it is recognized that the planning of tabulation forms is a task, the detailed execution of which must be carried on so as to fit each particular problem, there are certain general guiding principles which, if discussed here, may save the student or investigator much wasted time and effort.

Original and secondary tabulations. We speak of tabulation in general as "original" tabulation and as "secondary" tabulation. By original tabulation we shall mean the preparation of detailed tables on which are compiled the original data. By secondary tabulation we shall mean the preparation of tables which summarize the original data, and which permit comparisons of "groups" by means of "averages," measures of "variability," and measures of "relationship." The discussion of this chapter relates to the original tabulation of educational data. The complete treatment of secondary tabulation is included in Chapters IV to IX inclusive.

II. THE ORIGINAL TABULATION OF EDUCATIONAL DATA

I. HAND TABULATION

There are two important phases to the work of tabulation. The first has to do with the selection of the general scheme of tabulation, while the second deals with the method of tabulation.

We first face the question: What general scheme shall be used in compiling the original data — ruled cards, large ruled sheets, or ruled blank books? Two criteria control the selection of the general scheme:

(1) How many separate points are to be covered by the inquiry and how many cases are to be tabulated?

(2) Which method of tabulation is to be used, the "writing method" or the "checking method"?

Since the selection of the general scheme depends so completely on the adopted method of tabulation, that will be discussed next.

1. The method of tabulation

The writing method *vs.* the checking method. In compiling the original data of the inquiry, whether from question-blank returns or from original records, the investigator can adopt one of two procedures. He can write out the detailed data covering each point of his inquiry in the fashion indicated by Table 3. The data in the table are quoted from the illustrative tables in a study covering the social conditions and careers of more than five thousand teachers. It will be noted that the original data, compiled from the question-blanks, are written out in detail on this sheet. The only abbreviation of the data occurs in such questions as that covering "parental income," in which each of the

TABLE 3. EXTRACT FROM A TABLE OF ORIGINAL DATA ON THE CAREERS OF TEACHERS

Individual number	*Age*	*Parent's income*	*Number of months*	*Salary per month*	*Beginner's age*	*Rural school*	*Town school*	*City school*	*High school*	*Normal school*	*University*	*Nativity*	*Paternal language*	*Maternal language*	*Parent's occupation*	*Number of brothers and sisters*	*Family condition*	*Position*	*Board*
51	28	5	10	65	18	1	4	2	3	1	4	2	E	E	0	6	2	6	X
52	28	2	10	59	20	0	0	8	4	3	0	2	Sc	Sc	5	5	1	5	X
53	25	9	10	50	20	0	0	5	4	3	0	1	E	E	4	4	1	2	X
54	53	9	9	100	19	10	15	9	2	2	0	1	E	E	1	8	1	10	X
55	35	4	10	55	19	2	4	10	2	2	0	1	E	E	5	6	3	3	X
56	49	4	10	65	16	1	14	16	4	0	0	1	E	E	3	6	1	3	X
57	25	6	9	64	18	0	0	7	3	1	0	1	E	E	3	3	1	3	X
58	44	0	10	85	17	0	4	21	0	2	1	2	W	W	5	5	1	17	N
59	20	8	9	55	19	0	0	1	3	0	0	1	E	E	3	3	1	5	X
60	31	1	9	40	16	7	0	7	1	0	0	2	G	G	5	1	1	4	X

various intervals $250 or less, $250–$500, $500–$750, etc., is given a code number, and these numbers are tabulated, 5, 2, 9, 9, 4, etc. This is done, however, merely to save the time of writing the complete record for each teacher, and does not contribute at all to the more rapid summarization of the data later. In fact, in having to apply the code number to each case the tabulator is very likely handicapped in the rapidity and accuracy with which he compiles the data. It should be noted carefully that as a result of such a detailed original tabulation the original records are transcribed in full at a very considerable expense, but that no summarization has been done and none is possible on this table. The computation of averages and measures of variability and relationship for various comparable groups cannot be done without complete retabulation of the data. This brings us to a fundamental principle of tabulation: *the original tabulation should lead at once to group "totals," and to the rapid computation of the necessary statistical measures, averages, measures of variability, etc.* It is clear that the "writing method" of tabulation does not do this, and that for extensive investigations it is not an economical method.

The checking method. This brings up the checking method of tabulation, and we can illustrate its use by representing the same data given in Table 3. Its first distinctive feature is found in the form of the "heading" prepared for the tabulation. Now, instead of using a general blanket heading "age," for example, or "number of months" ("for which present contract is drawn") etc., we prepare a scheme of column headings, one column of the table being left for each possible reply, or perhaps for the smallest range covered by such replies. To illustrate, the column headed "age" in Table 3, now becomes a series of columns as in Table 4.

TABLE 4. PRESENT AGE OF TEACHERS

Teacher's Number	*17–19 years*	*20–22*	*23–25*	*26–28*	*29–31*	*32–34*	*35–37*	*38–40*	*41–43*	*44–46*	*47–49*	*50–52*	*53–55*	*65–68*	*Above 68*
51				x											
52				x											
53			x												
54													x		
55							x								
56											x				
57			x												
58										x					
59		x													
60					x										
Totals		1	2	2	1		1			1	1		1		

The records of "age" from Table 3 are retabulated by "checking" the appropriate column for each teacher. A second distinctive feature now stands out, — this method of tabulation at once permits grouping of data, and the immediate and easy compilation of totals, and of "averages" and measures of "variability." The student should be cautioned to classify his records carefully at the start so as to permit the tabulation of the data on a perfectly uniform group of individuals on one sheet or page. For example, the data given in Tables 3 and 4 should refer to teachers who are teaching under the same conditions, or who are from other standpoints perfectly comparable with each other. If this is done, as far as possible, the labor of retabulation in the subsequent statistical treatment of the data will be cut to a minimum.

To adopt the checking method on such points as "Age," in which many columns are needed, raises the question "How large shall the interval be made, — 1 year, 3 years, 5 years, or what?" Chapter IV discusses the statistical classification of data in great detail, and this question can best be answered for the reader by suggesting the reading of that chapter, with the subsequent rereading of this discussion. In that treatment a complete discussion of the size of the interval, its position, and best methods of marking limits, etc., are given.

To use the checking method, therefore, we must plan, at the start, a series of column headings sufficiently detailed to cover the range of possible replies on each point. The thought will arise immediately in the mind of the reader: "But the preparation of column headings is expensive, both of material and of the time of the tabulator." The first point is admittedly of not sufficient weight to demand consideration. The second is important, however. As the result of the detailed experience of the writer with both the writing and the checking methods, it can be said that the latter is by far the more economical in the long run. To offset the utilization of time in preparing column headings we have three distinct savings: (1) that due to *checking answers* instead of writing them out in detail; (2) that due to the possibility of totaling the data in each column *rapidly and accurately;* (3) the fact that *averages and other statistical measures can be computed for the various groups of data from the original record.* To these we should add that the checking method gives a more accurate perspective of the returns, permits better preliminary planning of the treatment of the data, and leads to a more adequate interpretation of results.

Schemes for tabulation. We said above that the selection of the scheme of tabulation depended not only on the method of tabulation, but also upon the number of points to be covered by the inquiry and the number of cases to be collected. In deciding on the scheme of tabulation we have a choice of the use of : (1) the ruled card; (2) the large ruled sheet; and (3) the ruled blankbook.

1. Use of the ruled card. It will be evident that the ruled card (regulation sizes, 4 by 6, 5 by 8, $8\frac{1}{2}$ by 11) is adapted to only the most restricted investigations, — those covering a comparatively small number of separate points and in which but few cases (perhaps 25 to 50) are to be collected. It has the advantage of facilitating manipulation and filing of the data. Such a scheme is excellently adapted to those compilations of data in which a single question may be put on a card, rulings being adapted in such a way as to give the data from each case on this particular point. This scheme is well adapted to the collection of data on various phases of city school administration by buildings, by kinds of schools, or by kinds of activities.

2. **Use of the large ruled sheet.** This is adapted to somewhat more extensive investigations, — those covering perhaps 50 to 100 points and as many cases. To the tabulation, for example, of the content of courses of study, or of the study of the content of textbooks, the large ruled sheet (19 by 24 inches is a standard size and easily manipulated) is well fitted. Its chief advantage lies in the clear perspective which it gives of all of the data covering a particular group of items or cases. It also permits easy secondary tabulation. It is used in large city systems, in many phases of the office tabulation of records; for example, in the standardizing of school supplies, both as to kind and amount, tabulation of "building" records, tabulation of bids, etc.

3. **Use of the ruled blankbook.** Nearly all educational investigations are extensive enough in number of points covered and in number of cases collected to demand tabulation of the original records in ruled blankbooks. A good rule is to use a book of standard size (say 8 by 10 inches) with cross-sectional ruling (to facilitate the non-uniform rulings which will be needed for data of the particular inquiry at hand) and including perhaps 60 to 100 pages. Thirty to forty cases can be tabulated on the length of the page. If

the checking method of tabulation is being used, the column headings should be arranged in the order of questions on the question-blank (if it is a question-blank inquiry), and the edge of the pages should be "cut-back" sufficiently to permit the use of the original list of names or numbers, written on the first page of the record. In this way, the entire record of an individual appears on the same line of the tabulation even though it may cover many pages in length. If the questions on the blank have been numbered consecutively as they should, these numbers could be used as column headings. There is almost no type of extensive investigation to which the ruled book is not well adapted, and in general it should have wide usage.

II. THE MECHANICAL TABULATION OF EDUCATIONAL STATISTICS

A recent development in statistical work. The discussion thus far has dealt with problems of school research which have implied the use of hand tabulation. For the tabulation and manipulation of the detailed educational and business records of a school system, the experience of school men is proving that electrical mechanical tabulation is both more economical and more efficient. Within the past few years, New York, Philadelphia, Rochester, Oakland (California), and other cities have adopted such methods and have proven their superiority to hand methods. To get the methods clearly before us, together with the consensus of practical judgment on their availability, quotations from recent discussions of the matter will be given.

Four methods. The following statement, by the Auditor of the Board of Education, New York City,[1] indicates four distinct methods of preparing statistical data: —

---

[1] Cook, H. R. M. "The Standardization of School Accounting and of School Statistics"; in *American School Board Journal*, June and July, 1913.

1. By means of the electrical tabulating and sorting machine and electrical battery adding machine, and by the use of perforated cards.

2. By means of cards of uniform size, on which are printed the statistical classifications, while the figures or amounts are inserted by hand. The margin of the card may be perforated by hand. This last process permits of a limited range of information being assembled. It also affords a means of assembling quickly all cards which relate to one or more items of classification. When assembling statistics from these cards, the use of the adding machine is advisable.

3. By the use of a columnar collateral ledger, exhibiting the various statistical classifications under which may be recorded the salient feature of the expenditure as shown by the voucher or by the voucher register.

4. By so planning the books of accounts as to include analysis columns in which should be entered, synchronously with the passage of a voucher, that particular statistical classification to which the expenditure may be applicable.

The first method is suitable either for large or for moderate-sized school systems, in fact, it may be profitably used anywhere, except in the case of the small rural organizations. In any city or town where the population exceeds 20,000 inhabitants, the installation of a statistical plant of this kind would be advantageous. Not only is it possible to make a complete distribution of school expenditures, but school facts of important character, both educational and physical, may be recorded with great speed, accuracy, and minuteness. A uniformly printed card, a few square inches in size, is susceptible of use for the purpose of recording thousands of facts of most varied nature. No matter how the cards may be fed through the machine, the sorting machine automatically separates each fact. The widest imaginable range of statistical information can be produced by the adoption of the first method. The system involves the compilation of a "code" in which each statistical point of information or fact is assigned to a number or combination of numbers. An illustration of the form of card used will be found among the diagrams. The cost of rental and operation of this type of statistical outfit in a small or moderate sized school system would be about the same as the salary of a clerk. In a large system it might reach the cost of two such clerks.

The second method is suitable for a system of any size and is very elastic, but it lacks the speed and wide range of the first-described method. It was actually and successfully employed for some years in one of the largest school systems in the world. It was only displaced because of the superiority of the first-described method, because the rental of a machine is cheaper than clerk hire. The cost of stationery is about the same. In a small school system the total cost would probably trend the other way, but not sufficiently far to make up for the extra efficiency and wide range of the mechanical device. An illustration suggestive of a suitable form of card will be found among the diagrams accompanying this treatise.

The third method represents a purely hand-made system, and is intended to operate in parallel with the regular books of account. The volume of the expenditures in the fund accounting will necessarily equal the volume of the statistical accounting between given points. This method permits of the preparation of data sufficient for the purposes of the standard blanks of the United States Bureau of Education, but it does not afford any very wide range of information which it might be desirable to collect for local purposes.

The fourth method is a modification of the third just-described method. It is suitable for school systems of a size which are required to present information for the purposes of the "abridged" standard blank adopted by the United States Bureau of Education.

All of the foregoing methods are practical. They have been tried and found to work successfully. They will furnish results within their limits and scope.

The Oakland, California, method. That the utilization of "mechanical tabulation" is not confined to the largest systems, but that it is efficient and economical in any city in which Tabulating Service Bureaus have been established, is shown by a recent report of Mr. Wilford E. Talbert, Director of Reference and Research, Oakland, California. After discussing the way in which the statistical reports of teachers, principals, and other employees are compiled by time-saving methods he says: —

Transferring reports to Hollerith cards. As soon as the teachers' reports are received in the Superintendent's office, the information they contain is punched onto Hollerith cards by a clerk who, be-

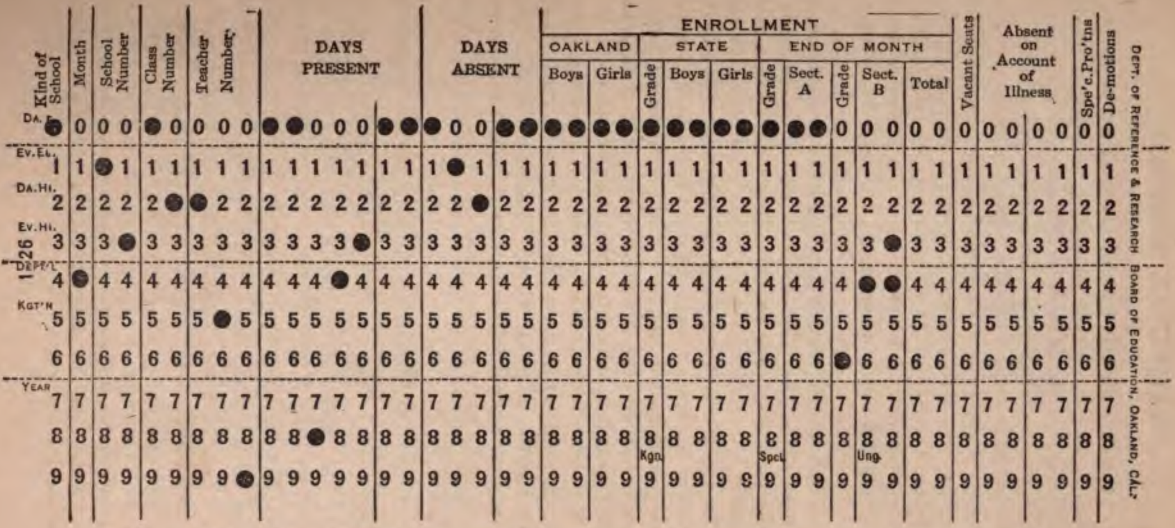


DIAGRAM 13. HOLLERITH TABULATING CARD USED IN OAKLAND, CALIFORNIA, SCHOOL SYSTEM

cause she specializes on this sort of work, is at least as apt to detect errors as any but the most careful principals. After the cards are punched they are all checked for accuracy by reading back the data to another clerk holding the original reports.

On the Hollerith cards (Diagram 13 reproduces an Oakland card), error has been carefully guarded against by color of cards, by clipping of corners, and by code numbers. Also an attempt has been made to foresee every possible kind of information that might ever be wanted from the reports for the given period. By the use of code numbers for sorting fields, this becomes very simple under the Hollerith system. In fact, someone has called the Hollerith cards "canned information," and, like canned goods, they are always on hand, they are compact, and their contents is always readily available. For example, it is possible in a few minutes on the sorting machine to take from the entire year's reports, the cards for special classes, for any given teacher, for any desired grade, for any teacher's register in the city (even though that register itself may have been burned), and all the data on these cards can be quickly tabulated in any desired way, even though none of this information is tabulated from month to month.

The work of tabulating results. As soon as all reports are received and all data have been transferred to the Hollerith cards, the latter are called for in the morning by the Tabulating Service Co., and the following four reports are returned by the same evening: —

1. Attendance and absence by schools and kinds of schools. (188 sums.)
2. Total enrollment by schools, by kinds of schools, and by departments of each school. (141 sums.)
3. Distribution of enrollment in non-departmental, and in departmental classes of the elementary schools, showing the number of classes of each size from the smallest to the largest, and giving the location by schools of all classes which are either exceptionally large or exceptionally small. (790 significant figures reported last month.)

These reports are all typed and arranged in such shape that this office can readily write in the names of schools and such averages, etc., as it is necessary to compute on the calculating machine. Even the typing is mechanically checked for error, so that we have absolutely reliable and unchangeable data as a basis for further computations.

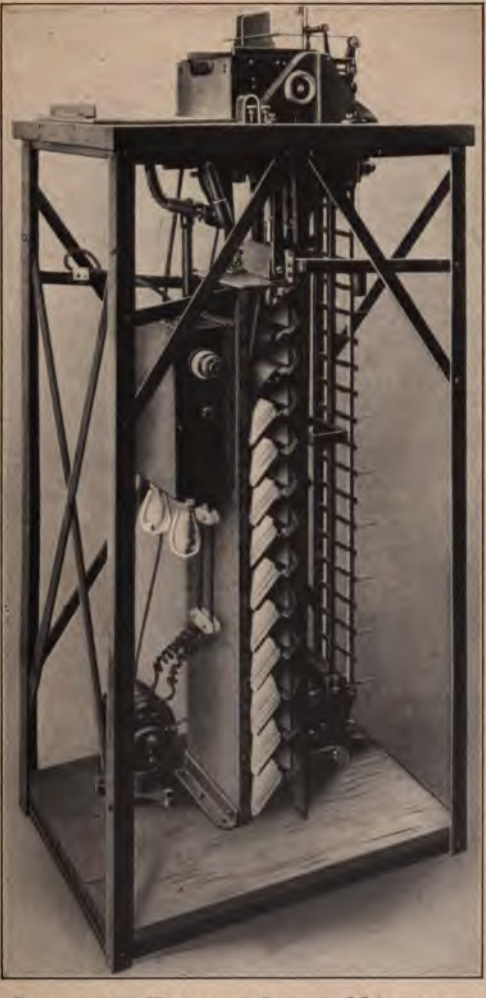


DIAGRAM 14. HOLLERITH SORTING MACHINE FOR CLASSIFYING SCHOOL STATISTICS

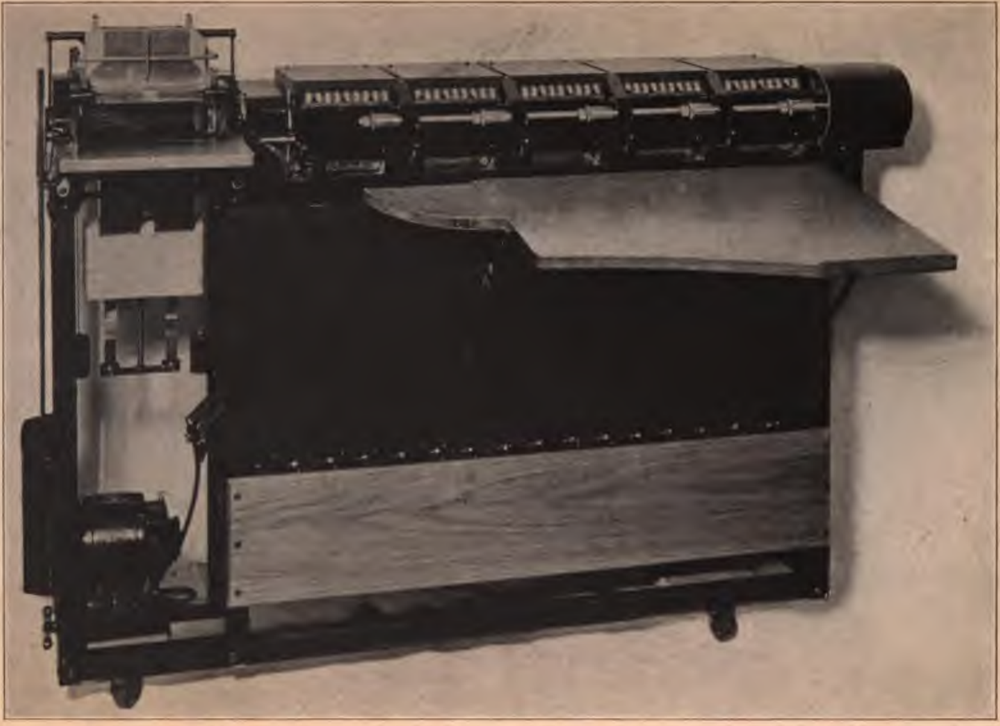


DIAGRAM 15. HOLLERITH TABULATING MACHINE USED IN RECORDING SCHOOL STATISTICS

Use of the plan at Rochester, New York. The statistical bureau of the school system of Rochester, New York, uses mechanical tabulation methods. Mr. J. S. Mullan, Secretary of the Board of Education, discusses the method in part as follows:[1]

Up to the present time, the analysis of school expenditures and the development of school statistics have been restricted because of cost and the time element, — that is, whether the information would be worth what it would cost, and whether it could be compiled in time for administrative and legislative use. With the adaptation of machinery to statistical purposes, we are entering upon a new era of statistical possibilities. With mechanical tabulation, cost and the time element are being reduced to the minimum. In fact, statistical analyses, heretofore prohibitive and practically impossible, are now being compiled, used, and demanded.

With mechanical tabulation, the bookkeeping division becomes a machine shop. The machinery consists of card-punching machines operated by hand (for individual cards and cards in gangs), a card-sorting machine [pictured in Diagram 14] operated by electricity, and a tabulating machine [pictured in Diagram 15], also operated electrically. The cards used in connection with the machines are somewhat larger than regular index cards. Upon the cards are printed what are technically known as "fields," each field representing an item of information. The field consists of vertical lines of varying distances apart, in which appear numerals, each field containing one or more perpendicular rows of numerals according to the requirements of each of the fields. The card which has been adopted for use in the accounting division of the Rochester school system shows the year and month; voucher number; vendor; school building; day, night, continuation or normal school; function; sub-function; educational subject; character of expenditure; quantity; unit of measure; commodity; class and number; price; amount; fund; and whether contract, open-market order, pay-roll, or miscellaneous expenditure.

Expressed in a numerical code, the information is punched on the cards by the operator striking keys which perforate the cards with

---

[1] "Mechanical Tabulation of School Financial Statistics"; in *Proceedings of the Fifth Annual Meeting of the National Association of School Accounting Officers*, p. 43.

small holes. Any data appearing on the requisition, invoice, payroll, or voucher can thus be transferred to the cards, after which the cards are ready for sorting and tabulation. It can be seen that once the cards are punched and checked with the original document, the period of detail checking is over. All the data punched on the cards are elemental. The total of the cards is the sum of the elements. Once punched and checked, the cards go to the sorting machine, where by electrical contact through the holes in the cards they are sorted into any pre-determined group; thence they go to the tabulating machine, where in the same way they are tabulated by groups and in total, the totals when obtained being entered on a prearranged form. The sorting and tabulating may be repeated until all the fields on the cards have been covered, the final totals of the various sortings being the automatic check. Furthermore, the punching of the cards and their tabulation are accomplished in a comparatively short period of time, so that any group result or combination of results is expeditiously produced, and at a minimum of cost.

Compare the possibilities of this procedure with distribution by hand posting, including the factor of possible clerical error, the difficulty of attempting to carry on more than one analysis at one and the same time, *i.e.*, functional service, amounts of compensation, quantities and prices of commodities, repairs, interest, refunds, bond payments, etc., and the confusion of thought in handling such a conglomerate, — and we begin to appreciate the possibilities of mechanical tabulation and its superlative advantages.

III. SECONDARY TABULATION

The future chapters. In Chapter II the initial steps in the study of an educational problem were shown to be the careful definition of the problem and the collection and original tabulation of the educational data. These were to be followed by the systematic classification of the data in the frequency distribution, and its summarization by means of various analytic and graphic methods. The discussion of the checking method pointed out that systematic planning of column headings for the original tables really amounted to the statistical classification of the facts. The principles and methods controlling this work are treated in detail in Chapter IV. The succeeding chapters, V to IX inclusive, take up the remaining steps in the statistical treatment of facts. In a fashion, they may all be called *Secondary Tabulation.* Chapter V presents the various methods of typifying data by "averages." Chapter VI shows how the data may be represented somewhat more completely by measures of "variability." Chapter VII discusses the methods of graphic representation of educational facts, and their connection with ideal frequency curves. Chapter VIII shows the application of such type curves to practical educational problems. In Chapter IX will be given a complete discussion of ways and means of determining the possibility and degree of relationship that exists between various aspects of school work.

CHAPTER IV

STATISTICAL CLASSIFICATION OF EDUCATIONAL DATA: THE FREQUENCY DISTRIBUTION

I. INTRODUCTORY

Statistics of attributes and variables. The study of the quantitative problems with which we deal in education reveals two principal statistical methods of treating the measurement of human traits: (1) the method of "attributes"; and (2) the method of "variables." The measurement of human traits may vary in refinement all the way from the mere counting of the presence or absence of a trait (treated by the method of attributes) to the rather minute quantitative measurement of the trait (better treated by the method of variables). The grouping of individuals according to the presence or absence of a trait may be illustrated by: the counting of the number of pupils in a class that have passed or not passed; the number that are of normal mentality, or are mentally deficient; the number that have light hair or dark hair, are tall or short, blind or seeing, sane or insane, and so on. The methods by which we would treat statistics collected in this way have been denoted by Yule, "THE STATISTICS OF ATTRIBUTES," and they are to be thought of as somewhat distinct from the methods of treating statistics collected by more refined methods of measurement. These latter, which are known as the "STATISTICS OF VARIABLES," imply that the specific magnitude of the trait has been measured with reference to a scale made up of known units. In general the statistics gathered in educational research are those of measurable traits, i.e., the statistics of variables. For example, we can measure, in a fairly refined way, the ability of pupils in arithmetic, or in algebra; the mental age of children; the cost of teaching various subjects of study; the retardation of pupils in the public school, and so on. We should have clearly in mind therefore that the statistical methods with which we treat one kind of statistics — those of attributes — may be different from those with which we treat the other kind — those of variables. The term variable as used in this book may be taken to mean a varying quantity or human trait, — for example, arithmetic ability, teaching skill, the height of men, etc. Thus, these traits are subject to statistical study by either mere enumeration or counting methods, or may be subject to fairly accurate measurement.

II. Classification of Statistical Data

Grouping of data into classes. Whatever may be the method by which, or the degree of refinement with which data are collected, when we turn to their organization so that we may interpret the situations that they represent, we face the problem of "grouping." Clear thinking about large numbers of facts necessitates the condensation and organization of the data in systematic form. That is, we are forced to group our data in "classes," and the statistical treatment of the data depends upon the determination of these "classes."

A statistical CLASS, whether of attributes or of variables, may be illustrated by Tables 5 and 6.[1] They picture the relation that exists, for example (Table 5) between the pedagogical standing and the mental standing of school children. To do this, the pedagogical ages of the children in question are grouped in three classes, — "retarded," "normal," and "advanced," — and the mental ages according to whether they are "retarded," "at level" (*i.e.*, normal) and "ad-

---

[1] From Stern's *Psychological Methods of Testing Intelligence*, pp. 59 and 61.

vanced." Corresponding to this same classification of mental age, Table 6 "groups" the pupils in three classes according to whether their school marks were poor, satisfactory, or good. Thus the 14 pupils "retarded" in both pedagogical age and mental age form a "class"; 16 that were "normal" in pedagogical age and "retarded" in mental age form another "class." Or, turning to the "total" columns, in the entire group of 101 there are found: a class of 24 pupils retarded, 65 pupils normal, and 12 pupils advanced. Because of the fact that refined quantitative methods were not employed in classifying the records we call these data, "STATISTICS OF ATTRIBUTES."

TABLE 5. RELATION OF PEDAGOGICAL AND MENTAL AGE *

Pedagogical Age	*Mental Age*			*Total*
	Retarded	*At level*	*Advanced*	
Retarded	14	9	1	24
Normal	16	33	16	65
Advanced	0	5	7	12
Total	30	47	24	101

* Binet.

TABLE 6. RELATION OF MENTAL AGE AND SCHOOL MARKS †

School Marks	*Mental Age*			*Total*
	Retarded	*At level*	*Advanced*	
Poor	29	17	0	46
Satisfactory	26	79	21	126
Good	0	13	31	44
Total	55	109	52	216

† Bobertag.

Suppose, however, that the standing of the 216 pupils represented in Table 7, instead of being grouped as poor, satisfactory, and good, had been given in terms of numerical marks on a 100 per cent scale, say, — 87, 82, 54, 76, 91, etc. It will be clear that the grouping of these data now necessitates setting definite numerical limits to the classes in which the various measures (individual marks), are going to fall. Now, instead of being called poor, satisfactory, good, the marks will be found to fall within some definite interval of the scale, 85.0 to 89.99; 80.0 to 84.99; 50.0 to 54.99; 75.0 to 79.99; 90.0 to 94.99, etc. Our data thus illustrate again the STATISTICS OF VARIABLES, and point out the differences between the method of treating such measures and the ATTRIBUTES represented in Stern's tables.

Distribution on scales. This discussion of the grouping of measures has made use of several important concepts, which must be clearly grasped by the student. Fundamental to the practice of measurement are the concepts of SCALE and UNIT. We shall think always of mental and social measurements as distributed over a "scale" — *i.e.*, a linear distance or a difference in numerical magnitude which will represent or stand for the magnitude of the measures in question.

For example, the ability of a group of children in handwriting may vary in magnitude, let us say, from 40 to 75, when measured on a total scale of "handwriting merit," such as is given in the *Scale for Measuring Handwriting*, devised by Dr. L. P. Ayres, from 20 to 90.

Scholastic abilities are measured, very generally, by the percentile marks of teachers, which are taken to represent the relative position of pupils on a one hundred per cent scale. The per-pupil costs of teaching the various high-school subjects may be pictured clearly as distributed over a "cost-scale." The scale may be pictured in numerical or graphic

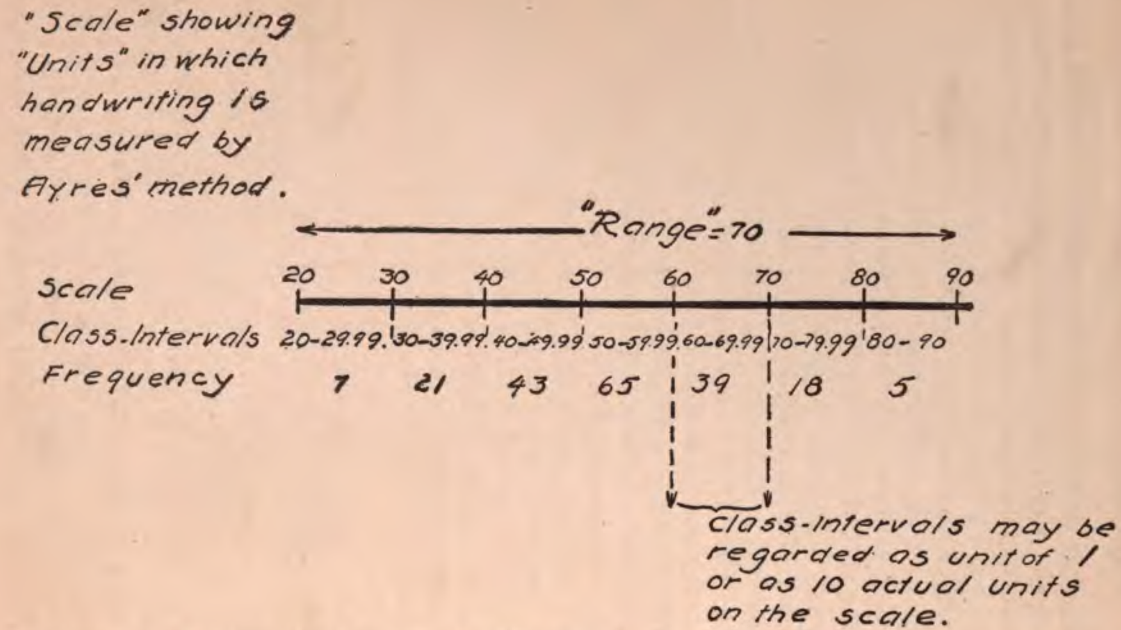


DIAGRAM 16. TO ILLUSTRATE USE OF "SCALE," "UNIT," "CLASS-INTERVAL," AND "FREQUENCY DISTRIBUTION"

terms. Let us illustrate these points by graphic illustrations. The student will be aided in grasping the reasons for certain steps in statistical computation if he will always

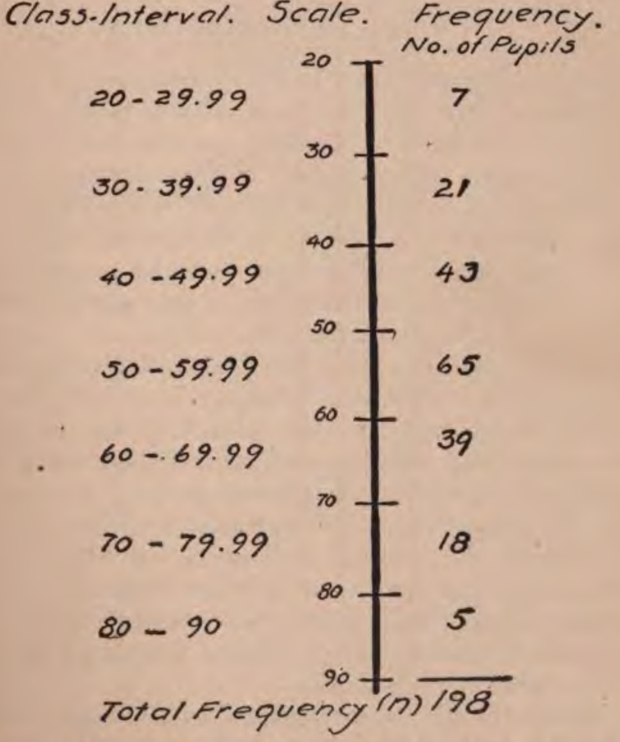


DIAGRAM 17. TO ILLUSTRATE USE OF "SCALE," "UNIT," "CLASS-INTERVAL," AND "FREQUENCY DISTRIBUTION"

supplement his numerical thinking about the "scale" with a graphic picture of it. For example, Diagrams 16 and 17 give a numerical representation of the handwriting scores of 198 pupils grouped in various CLASS INTERVALS along a SCALE, whose RANGE (the distance from the smallest measure to the largest measure) extends from 20 per cent to 90 per cent.

III. Classes and Class Limits

Manifold classification. In distinction from the rough grouping of attributes illustrated above, this numerical classification of measures is called "manifold-classification." The student should be cautioned that the classes should be clearly marked off from each other by definite numerical limits, 50.0–54.99; 55.0–59.99; or 47.5–52.49; 52.5–57.49, etc., if the class-interval is to contain, for example, five units. There are three different ways in which the limits may be set to the intervals on the scale:

The first method of setting class limits is to give the limits themselves, as: 5.0–9.99; 10.0–14.99; 15.0–19.99, etc., as is given in Diagram 17. The student, in beginning the tabulation of frequency distributions, is advised to make use of this definite method, clearly distinguishing the position of the intervals. Especially is it a helpful devise in increasing the accuracy of tabulation. The use of the method 5–10; 10–15; 15–20, etc., leads to many errors in tabulation. The routine statistical work should be safeguarded at every possible step. Clear marking off of class-intervals will tend to reduce errors in this particular.

The second method of setting class limits is to express the interval in terms of the mid-value of the class-interval; for example: 7.5; 12.5; 17.5. From the standpoint of accuracy in tabulating the frequencies this is a very poor method, and leads to many errors in tabulation.

The third method is to state the interval in words in the form, "5 and less than 10"; "10 and less than 15"; "15 and less than 20," etc. As cautioned above, the use of the same numbers in expressing the numerical limits of class-intervals, 10, 15, 20, etc., and the complication of the word-heading leads to error. It should be clear that, at least for the novice in statistical work, class-intervals should be defined very carefully. Students consistently make more mistakes in the routine tabulation of measures than in the computation of means, measures of variability, etc., after the data have been arranged.

IV. The Frequency Distribution: The Steps in its Construction

Arranging a frequency distribution. The grouping or classifying of measures consists, therefore, (1) in noting the length of the range, i.e., the distance between the largest and the smallest measures; (2) in deciding on the number of class-intervals (or, the size of *a* class-interval) into which you are to divide the total range of the measures; (3) setting the *position* of the class-intervals (*i.e.*, determining the specific class-limits); and (4) tabulating the FREQUENCY of occurrence of the measures in each of the class-intervals. The result of such grouping of measures is called a "FREQUENCY DISTRIBUTION," and is made up of two columns of figures, first a serial list of the "CLASS-INTERVALS," arranged preferably with the smaller measures at the lower end of the scale; second, a column of "frequencies," which gives the number of measures tabulated in each class interval. Tables 7 and 8[1] give illustrations of the frequency distribution as it is used in the study of educational problems, and which make use of the method of defining class-limits very carefully.

---

[1] Judd, C. H., and Parker, S. C., *Problems Involved in Standardizing State Normal Schools*, pp. 17, 18, 19. Bulletin no. 12, U.S. Bureau of Education. (1916.)

TABLE 7. ADVANCED DEGREES HELD BY MEMBERS OF NORMAL SCHOOL FACULTIES

Percentage of faculty	*Colleges and universities*		*Normal schools*	
	*Ph.D.**	*Master**	*Ph.D.†*	*Master†*
0 to 9	2	1	22	3
10 to 19	11	1	8	6
20 to 29	16	1	2	5
30 to 39	13	2	..	7
40 to 49	6	7	..	6
50 to 59	10	8	..	4
60 to 69	2	11	..	0
70 to 79	3	11	..	1
80 to 89	..	15	..	..
90 to 100	..	6	..	..
Total.............	63	63	32	32

* Nine not reporting. † Three not reporting.

TABLE 8. AVERAGE SALARIES IN NORTH CENTRAL COLLEGES AND NORMAL SCHOOLS

Salaries	*Universities and colleges*	*Normal schools*	*Salaries*	*Universities and colleges*	*Normal schools*
$900 to $999	..	1	$1600 to $1699	9	3
1000 to 1099	3	..	1700 to 1799	9	3
1100 to 1199	4	1	1800 to 1899	2	5
1200 to 1299	8	2	1900 to 1999	5	1
1300 to 1399	6	1	2000 to 2099	1	..
1400 to 1499	7	5	2100 and over	2	3
1500 to 1599	6	3	No information	10	7
Total	34	13	Total	38	22

The first step in constructing the frequency distribution. To make clear the construction of the frequency distribution let us work through a problem with the following illustrative data. Table 9 gives the "original measures," — in this case, the marks given to 123 pupils in English. Running down each of the columns we note that the lowest mark given was 20; the highest, 95. Thus the range is 75. In the treatment of these data our aim is to classify them in such a way that, for example, an "average," computed for the data in the classified or condensed form, will be very closely the same as the "true average," which would be computed from the entire list of the original measures themselves.

TABLE 9. CLASS MARKS GIVEN TO 123 HIGH-SCHOOL PUPILS IN ENGLISH

80	57	45	74	95	80	73	87	59	80	57	52
75	75	63	75	84	50	77	76	63	90	79	80
58	71	60	85	76	76	72	73	56	75	84	80
87	85	69	85	40	66	78	79	73	86	88	75
80	79	80	60	87	80	78	82	52	75	67	80
77	80	66	74	73	79	60	66	57	74	76	70
55	87	87	72	73	68	87	81	60	75	35	73
75	67	78	86	73	79	40	82	55	65	80	86
79	65	73	56	71	73	80	67	78	62	79	79
81	77	82	78	93	78	70	72	79	45	81	75
20	80	30									

The second step in constructing the frequency distribution. This is: deciding on the number of class-intervals into which the range shall be divided; *i.e.*, how many units on the scale shall be included in one class-interval. Two questions have to be answered: —

(1) How large may the class-interval be made and still give reasonably small errors in the computation of "averages," etc. The larger we make the interval — that is, the more greatly measures are condensed — the more do we cut down our labor of arithmetic computation. In an extensive investigation, which includes many frequency distributions made up from data that show similar characteris-

tics as to variation, it may be feasible to take the time to group the data in several different frequency-distributions, computing, say, some average value for each. If the student does so he will note that as he makes the size of class-interval smaller there will be an "optimum size" beyond which further reduction will not give an increase of accuracy of the average. In most educational investigations, however, an empirical rule can be given to guide the student in his work. In general, when the units of the scale covered by the range are as few as 10, 15, or even 20, nothing is to be gained by grouping the data in fewer classes, and we may let each unit represent a class-interval. For example, in the problem given in Table 18 there are 12 different unit costs of teaching English, the frequency of the occurrence of each of which is given for 148 Kansas cities. The true mean may be rapidly computed without grouping. On the other hand, the 123 class marks given in the foregoing problem cover a range of 75 units, and obviously must be grouped. A practical rule is to condense to not more than 20 intervals, and to choose a size that gives ease of tabulation. In this case class-intervals of 5 units convert a range of 75 units into 15 class-intervals, a good working number.

(2) In what ways are the measures concentrated around certain average values? For example, are most of the marks in the illustrative example grouped in the middle of the scale, with about the same number of measures on each side (that is, do they form a fairly "symmetrical" distribution), or are they widely scattered over the scale, each value occurring only a few times? This question can be answered roughly by careful inspection of the lists of original measures. If such inspection leads to the conclusion that the measures are fairly well concentrated, or are symmetrically dispersed over the scale, the particular method of grouping will not cause a fluctuation in the value of the "average" or the

"measure of variability" that is computed from the frequency distribution.

Fundamental assumption underlying grouping. There is one fundamental assumption that we make in all "grouping" of measures in a frequency distribution, namely, that all the values in any class-interval are concentrated at the mid-point of the interval, and may be represented by the value of this mid-point. For example, if the data of Table 9 were grouped in class-intervals of 5 per cent, as in 73.0–77.99; 78.0–82.99; etc., then the values of 74, 73, 75, 76, 77 all fall within the interval 73.0–77.99, and for all practical purposes are each assumed to be equal to the mid-value, 75.5. It will be clear that, with very unsymmetrical distributions, the assumption is untenable as large errors of computation come about. For example in the cost problem on page 116, grouping the original distribution in class intervals of 2 makes at the low end of the range a very material error in the first interval, one city actually having a per-pupil recitation cost of one cent, and 26 cities a cost of two cents, the "grouping" causing us to assume that 27 cities each have a cost of 1.5 cents. The error in computing the "average," due to this assumption, is partly compensated for, however, by the next interval in which are combined 46 measures at three cents, and 26 measures at four cents, offsetting in part the "skewing" of the average toward the low end of the scale. In some educational investigations the data are either so scattered, or are concentrated unsymmetrically, — more heavily at one end of the scale, — that it is necessary to be cautious about grouping. It should be pointed out that with most educational measurements the data are concentrated fairly near the middle of the scale, and tend to be fairly symmetrical. This is a fortunate condition, and makes relatively easy for the student the problem of grouping his data. In general he may accept it as a rule, for guiding the preparation of frequency distributions, that he should get a working number of class-intervals, say from 10 to 20, but at the same time should make the interval as small as is necessary to reveal any particularly predominant points on the scale.

Summary as to class-intervals. Summing up the foregoing statements on the question of deciding the number of class-intervals, we see that the class-interval must be made as large as is possible, and at the same time give relatively slight error in computation from the frequency distribution; that the intervals should not exceed approximately 20 in number, or, in general, be less than 10; that the grouping can be done much more completely if the measures are concentrated fairly near the middle of the range, and are distributed in a somewhat symmetrical manner on both sides of this general point of concentration; that in all grouping we make the very important assumption that all measures in a class-interval are grouped at the mid-point of the interval, and are equal to it in value, and that this assumption is pertinent to the determination of the size of the interval, the larger and more unsymmetrical the distribution of measures in the interval the greater the error made in making the assumption.

Third step in constructing the frequency distribution — determining the position of the class-intervals. In dividing up the range into class-intervals we are forced to decide at what digits to set the numerical limits of the intervals, — 50.0, 55.0, or 47.5, 52.5, 57.5, or 53, 58, 63, etc. Two criteria control this decision: First, the interval should be set at such points on the scale as will lead to the greatest ease and accuracy of tabulation. The experience of the writer and his students leads to the belief that to satisfy this criterion, intervals should not only start and stop with digits, but should make use of the basic tens system wherever pos-

Table 10

Classification I.

123 original measures grouped in class intervals of 5 units each,
so arranged as to make the values of interval integers.

Class Interval	Value of Mid-point	. Results of First Step of the Tabulating Measures	Frequency f
95.0-100	97.5	/	1
90.0-94.99	92.5	//	2
85.0-89.99	87.5	𝍸 𝍸 ///	13
80.0-84.99	82.5	𝍸 𝍸 𝍸 𝍸 /	21
75.0-79.99	77.5	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 /	31
70.0-74.99	72.5	𝍸 𝍸 𝍸 ////	19
65.0-69.99	67.5	𝍸 𝍸	10
60.0-64.99	62.5	𝍸 //	7
55.0-59.99	57.5	𝍸 ////	9
50.0-54.99	52.5	///	3
45.0-49.99	47.5	//	2
40.0-44.99	42.5	//	2
35.0-39.99	37.5	/	1
30.0-34.99	32.5	/	1
25.0-29.99	27.5		
20.0-24.99	22.5	/	1
Total			123

TABLE 10 (*continued*)

Classification II.

Class Intervals of 5 units each; mid points are integer			
Class Interval	Mid-point	Results of Checking	Frequency f
92.5-97.49	95	//	2
87.5-92.49	90	//	2
82.5-87.49	85	𝍸 𝍸 ////	14
77.5-82.49	80	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 ////	34
72.5-77.49	75	𝍸 𝍸 𝍸 𝍸 𝍸 ///	28
67.5-72.49	70	𝍸 ////	9
62.5-67.49	65	𝍸 𝍸	10
57.5-62.49	60	𝍸 ///	8
52.5-57.49	55	𝍸 /	6
47.5-52.49	50	///	3
42.5-47.49	45	//	2
37.5-42.49	40	//	2
32.5-37.49	35	/	1
27.5-32.49	30	/	1
22.5-27.49	25		
17.5-22.49	20	/	
Total			123

Table 10 (*continued*)

Classification III.

Class Intervals of 3 units each.			
Class Interval	Value of Mid-point	Results of Checking	Frequency f
94.0-96.99	95.5	/	1
91.0-93.99	92.5	/	1
88.0-90.99	89.5	//	2
85.0-87.99	86.5	𝍸 𝍸 //	12
82.0-84.99	83.5	𝍸	5
79.0-81.99	80.5	𝍸 𝍸 𝍸 𝍸 𝍸	25
76.0-78.99	77.5	𝍸 𝍸 ///	13
73.0-75.99	74.5	𝍸 𝍸 𝍸 𝍸 /	21
70.0-72.99	71.5	𝍸 //	7
67.0-69.99	68.5	𝍸	5
64.0-66.99	65.5	𝍸	5
61.0-63.99	62.5	///	3
58.0-60.99	59.5	𝍸 /	6
55.0-57.99	56.5	𝍸 //	7
52.0-54.99	53.5	//	2
49.0-51.99	50.5	/	1
46.0-48.99	47.5		0
43.0-45.99	44.5	//	2
40.0-42.99	41.5	//	2
37.0-39.99	38.5		0
34.0-36.99	35.5	/	1
31.0-33.99	32.5		0
28.0-30.99	29.5	/	1
25.0-27.99	26.5	/	0
22.0-24.99	23.5		0
19.0-21.99	20.5	/	1
Total			123

Table 10 (*continued*)

Classification IV.

Class Intervals of 10 units each; mid point of each interval an integer

Class Interval	Value of Mid-point	Results of Checking	Frequency f
90.0-100	95	///	3
80.0-89.99	85	𝍸 𝍸 𝍸 //// 𝍸 𝍸 𝍸	34
70.0-79.99	75	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸	50
60.0-69.99	65	𝍸 𝍸 𝍸 //	17
50.0-59.99	55	𝍸 𝍸 //	12
40.0-59.99	45	////	4
30.0-39.99	35	//	2
20.0-29.99	25	/	1
Total			123

sible. Thus the measures given in Table 10, Classifications I and II, make use of this method, — 50.0–54.99; 55.0–59.99; 60.0–64.99, etc. The second criterion has to do with the later manipulation of the measures in the frequency distributions, — such as is required in the working of the weighted arithmetic mean. Such computation requires the multiplication of the frequencies by the mid-points of the class-intervals. To cut down the arithmetic labor involved in this process would seem to demand that the mid-point be an integer, — for example, 55, 60, 65, 70, etc. Classifications I, II, III, and IV of the data in Table 10 illustrate the differences in the computation with the mid-points at integral and decimal points.

The later discussion of the computation of averages and variability shows, however, that the actual multiplication may all be reduced to mental processes (by the use of short methods). For this reason the second criterion should not hold in deciding on the position of class-intervals. It is the writer's judgment that accuracy and rapidity of tabulation should guide the student, and cause him to use that classification of limits for his intervals that bring about the most rapid and most accurate tabulation. It is recommended that for distributions covering a large portion of the percentile range, intervals of 5 be used, and that their limits be set at 20.0, 25.0, 30.0, 35.0, etc.

V. THE GRAPHIC REPRESENTATION OF EDUCATIONAL DATA

Importance of graphic representation. The fundamental aim of all statistical organization of educational data is to secure clear interpretation of the situation represented by the data. The numerical classification of large numbers of facts in the frequency distribution is certainly the first important step in condensing the original measures so that the mind can deal clearly with them. It will be shown in the next two chapters that there are two major numerical methods of further condensing the material, — the method of "averages," and the method of "variability." Each of the methods condenses the facts of the frequency distribution into a single number, and aids materially in the interpretation of the data. But thorough use can be made of such measures only by the most experienced manipulator of statistical methods. The student needs still more concrete methods of representing facts. Probably the greatest aid to sound interpretation of statistical data will come from the *graphic representation of the facts in question.* At this point, then, it will be well to take up a brief discussion of the plotting of frequency distributions.

Representing a frequency distribution. There are two principal methods of representing a frequency distribution by a graph: (1) that which gives the FREQUENCY POLYGON; and (2) that which gives the HISTOGRAM or COLUMN DIAGRAM. The two methods are illustrated by Diagrams 20 and 21, which graphically represent the data of Table 10, in three different classifications. In both types *the horizontal base line represents the scale* along which the class-intervals of the frequency distribution are laid off. The class-intervals are laid off on this scale by making use of the largest "unit" that the width of the paper will permit. *The vertical lines represent the number of measures* found to fall in a particular class-interval or at a particular point on the scale.

General directions for plotting. All graphing is done on two basic lines or axes. Using our established notation we may call these OX and OY. Keeping the accepted algebraic methods of graphing we shall lay off all units on the horizontal scale from left to right, and all units on the vertical scale from bottom to top. Doing this, as in Diagram 18, the steps in making the frequency polygon are these: —

1. Note the numerical amount of *the range of the frequency distribution.*

2. Lay off the units of the frequency distribution on the base line OX. Make the units as large as possible and yet get all of the distribution on one piece of paper. Obviously the selection of units must be left to the judgment of the draftsman. Mark clearly the limits of the class-intervals on the base line.

3. At the mid-point of each class-interval draw a vertical line, *the length of which represents, to any selected scale, the number of measures* that have been found to fall within that class-interval. If your data are definite integral records, varying by units of one each, such as the number of problems solved by large numbers of pupils in arithmetic, — say, 10 solving 5, 14 solving 6, 72 solving 7, 158 solving 8, 49 solving 9, 10 solving 10, etc., — then draw the vertical lines representing the number of individuals at these definite unit points, 5, 6, 7, 8, 9, 10, etc. No grouping of records is done, and no assumption is made that the measures are concentrated at the mid-point of the class-interval.

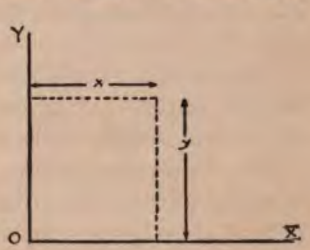


DIAGRAM 18. TO ILLUSTRATE USE OF COÖRDINATE AXES X AND Y

All measures plotted on OX are called "x"; all on OY, "y."

Size of unit. The selection of the size of the "unit" in laying off the number of measures on the vertical lines is arbitrary. Two principles of construction should control it however: (1) The units should be made large enough that the whole distribution may be pictured on one graph — the size of the paper chosen will determine this point. (2) The units must be large enough to make very clear the characteristic features of the distribution. This means that the horizontal and vertical scales shall be so taken that the polygon is sufficiently "steep" to indicate distinct changes in the distribution of the data. Especially is this true of graphs which picture rates of increase, in which case we should avoid using a small scale, which will result in a very "flat" polygon.

The student should be directed to indicate very clearly on the graph: (1) the limits of the class-intervals; (2) the distribution of the units along the vertical axis OY.

The frequency polygon. Diagrams 19, 20, and 21 illustrate the plotting of the frequency-distribution for two kinds of records: (1) ungrouped measures expressed in integral units; (2) measures grouped in class-intervals. The measures for the first illustration are arranged in the frequency distribution, shown in Table 11. Such a table is then plotted as is shown in Diagram 19.

TABLE 11. NUMBER OF FACTORING PROBLEMS SOLVED CORRECTLY BY 137 PUPILS IN FIRST-YEAR ALGEBRA

No. of problems	*No. of pupils*
13	1
12	3
11	8
10	14
9	29
8	35
7	21
6	16
5	5
4	3
3	1
2	1
	137

Diagram 20 illustrates the plotting of measures which have been grouped in the frequency polygon. The student should be reminded again of the fundamental assumption underlying this method, namely: the values of all measures in the class-interval are assumed to be equal to the mid-value of the interval, and *in plotting are actually concentrated* at this mid-value. An important corollary to the above statement, then, is this: the total number of measures in the frequency distribution is equal, to scale, to the total length of all the vertical distances laid off above the mid-points of the class-intervals.

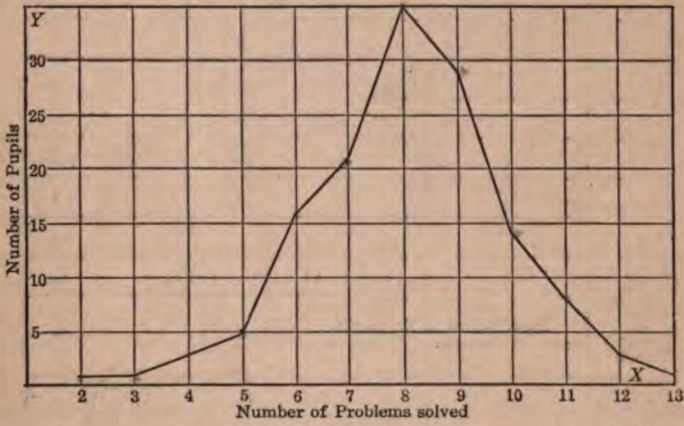


DIAGRAM 19. FREQUENCY POLYGON REPRESENTING INTEGRAL MEASURES

Based on Table 11, showing the number of factoring problems solved correctly by 137 pupils in first-year algebra.

The histogram, or column diagram. Thus the procedure stated above for the plotting of frequency polygons represents the frequency distributions by the length of vertical lines erected at the mid-points of class-intervals. Another specific method of graphically representing the distribution of measures over a scale is to assume that the measures may be represented by the *area of rectangles*, constructed with

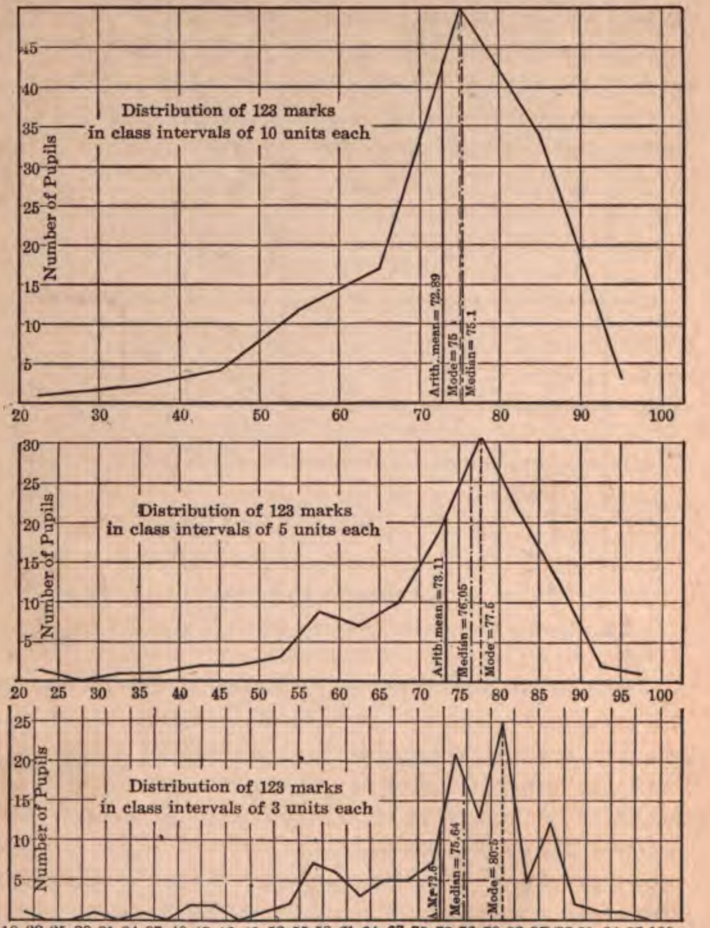


DIAGRAM 20. TO ILLUSTRATE THE PLOTTING OF THE "FREQUENCY POLYGON" FOR A GROUPED DISTRIBUTION

the base of the rectangle equal to the length of the class-interval, and the altitude equal (to the chosen scale) to the number of measures in that class-interval. (Diagram 21.) It will be clear that *such plotting of measures makes the definite assumption that measures are distributed uniformly throughout the interval.* This assumption is to be contrasted with the one made in the case of the frequency polygon, — namely, that all measures in the class-interval are concentrated at the mid-point of the interval.

Plotting of this kind has to do with a definite and generally fairly *small* number of measurements which have been made, in educational research, with rather rough measuring instruments. With the development of very refined methods of measuring, and the collection of a large number of measures, the measures would be found to vary from each other by very small amounts. Furthermore, human measurements, when compiled in large numbers, point to the fact that the numbers of measures at consecutive points on the scale are closely the same. That is, as we increase the accuracy of measurement and the number of observations, the mid-points of our class-intervals move more and more closely together. Furthermore, the tops of the ordinates erected at these points tend to form a continuous curve, instead of a polygon of broken lines. This curve we speak of then, as a FREQUENCY CURVE, and the *total area between the curve and the base line represents the total number of measures.* This is important for the student to hold in mind in connection with the later graphic treatment of measures.

It will be evident that the area under the frequency polygon represents very inadequately the number of measures in the distribution, in those cases in which the number is small and the range is relatively large. In such cases it is suggested that the column diagram be drawn, as typifying more clearly, by its area, the true status of the measures.

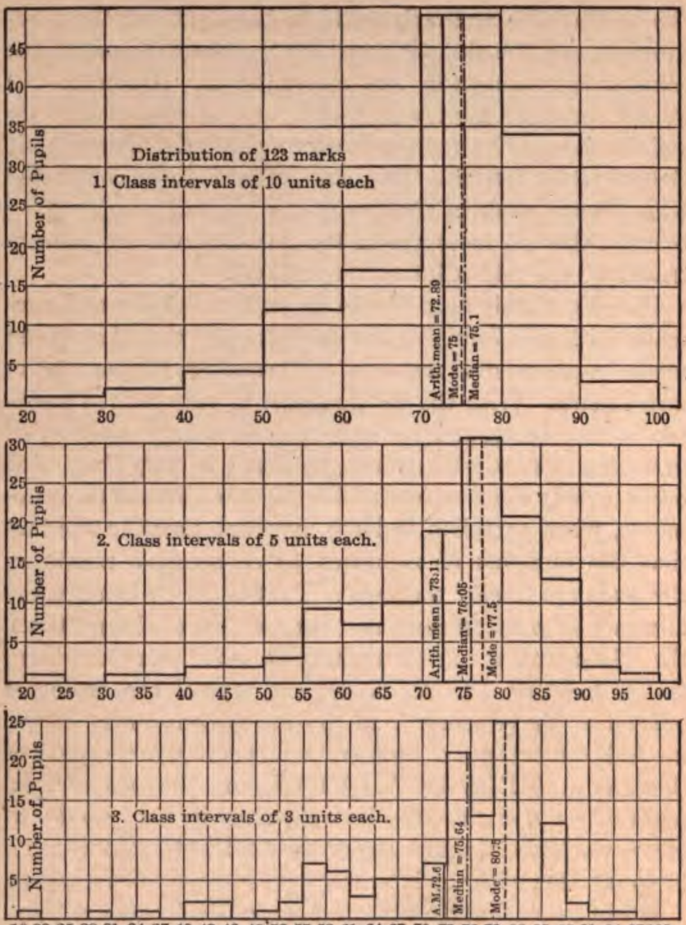


Diagram 21. To illustrate the Plotting of a "Column Diagram"

ILLUSTRATIVE PROBLEMS *

1. Tabulate the following series of measures in 3 frequency distributions, using class-intervals of 5, 10, and 15 units each respectively.

Distribution of Number of Pupils Taught by One Teacher in Science in 125 Small Cities

147	66	70	61	126	63	85	54	92	73	96	44	53
45	95	87	48	98	76	75	52	50	115	36	78	51
58	52	77	45	68	37	62	53	60	38	109	40	41
75	90	93	57	94	64	52	44	84	97	94	50	46
85	71	46	73	67	77	47	54	47	93	102	60	54
152	151	108	81	80	86	50	117	78	62	74	55	
72	86	93	143	78	92	41	91	87	82	89	52	
145	76	79	50	35	76	56	105	48	88	61	40	
121	71	39	132	88	101	70	95	91	71	64	50	
107	91	74	59	77	63	62	72	82	111	83	58	

2. Tabulate each of the following series of measures in a frequency distribution. Use your own judgment concerning the best size and position of class-interval.

Average Annual Cost per Pupil for Stationery in 122 St. Louis Schools, 1910–11 and 1914–15†

Series I (1910–11)					Series II (1914–15)					
2.78	.61	.29	.80	.53	1.80	.50	.38	.61	.46	.18
.88	.64	.54	.59	.49	1.20	.58	.37	.44	.55	.50
1.61	.41	.58	.51	.53	1.38	.38	.43	.51	.57	.45
1.15	.74	.39	.52	.51	1.58	.54	.41	.37	.45	.15
1.50	.71	.53	.60	.47	1.61	.58	.45	.51	.53	.26
1.59	.66	.72	.70	.58	1.78	.46	.39	.40	.47	.20
.43	.54	.66	.41	.43	.39	.40	.41	.56	.47	.28
.48	.49	.50	.52	.51	.41	.44	.45	.56	.47	.26
.54	.61	.66	.53	.50	.45	.33	.52	.48	.58	.55
.50	.45	.67	.57	.53	.52	.45	.51	.43	.49	.16
.59	.62	.46	.45	.47	.62	.48	.35	.56	.36	.33
.50	.48	.38	.43	.45	.46	.68	.54	.38	.57	.50
.56	.56	.58	.62	.25	.54	.39	.58	.37	.41	
.57	.32	.64	.91	.32	.47	.53	.44	.26	.54	
.62	.71	.38	.44	.33	.55	.35	.57	.59	.48	
.50	.50	.56	.43	.39	.40	.58	.43	.62	.48	
.71	.51	.60	.53	.18	.43	.52	.31	.36	.44	
.34	.65	.72	.49	.14	.51	.40	.34	.59	.49	
.56	.63	.57	1.04	.24	.53	.57	.52	.42	.72	
.44	.42	.58	.57	.71	.41	.55	.58	.68	.57	
.30	.56	.50	.59	.51	.28	.31	.43	.43	.84	
.81	.34	.44	.45		.47	.62	.65	.50	1.66	

---

* These illustrative problems are quoted from Rugg, H. O., *Illustrative Problems in Educational Statistics*, published by the author to accompany this text. (University of Chicago, 1917.)

† Data from *Annual Reports*, Board of Education, St. Louis, Missouri, 1914–15.

3. Plot a *frequency polygon* for each of the three distributions of problem No. 1. Select such a scale on X and Y that you can plot the three graphs one above the other on one cross-section sheet. Place the graphs so that corresponding points on the scales of the three distributions will fall on the same vertical line.

4. For the data given in each series in problem No. 2, plot a *column diagram*. Select such a scale on X and Y that you can plot the three graphs one above the other on one cross-section sheet. Place the graphs so that corresponding points on the scales of the three distributions will fall on the same vertical line.

CHAPTER V

THE METHOD OF AVERAGES

THE FIRST METHOD OF DESCRIBING A FREQUENCY DISTRIBUTION

1. *General statement of methods of describing a frequency-distribution*

Measures of condensation and organization. Having discussed the method of organizing material in the form of a frequency distribution, we are now prepared to take up the consideration of methods of statistically treating the distribution. It seems clear that the organization of the material in serial class arrangement, as in a frequency distribution, is but a preliminary step to the definite quantitative treatment of the material itself. The frequency distribution with its accompanying diagrams may represent adequately the status of the numerical data. It does not, however, enable definite comparison of its central tendency (*e.g.*, the "average") with that of the typical status (the "average") of other distributions. To make these comparisons, to be able to portray the typical numerical situation concisely and completely, *we need measures of condensation and organization*. There are three principal methods of typifying frequency distributions that will aid in comparison.

Central tendency and variability. The first method is the method of averages or of "central tendency"; *i.e.*, the method that shows how distributions differ in *position*, as shown by the size of the measure around which the measures largely cluster. The second method is, *method of variability; i.e.*, the method that indicates the way in which the separate measures of two distributions, "spread," or "fluctuate," around the "average." It is clearly not sufficient to be able to compare the status of two distributions by stating their average value. The average value may be and often is a deceptive value for use in comparative work. In fact, any one statistical measure will probably be an inadequate means of fully describing any group of data. This is clearly shown by the frequency curves drawn in Diagram 22, in which the average value of the two distributions is identical, but in which one distribution is twice as *variable* as the other. In this illustrative case it would be quite incorrect to infer from the identity of the average standing of the two classes that the distributions of abilities are equal. In this case at least we need a measure which will enable us to compare *the variation of ability in the two classes about the average ability of each.*

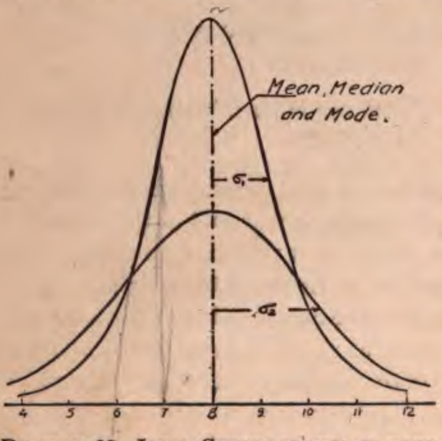


DIAGRAM 22. IDEAL CURVES DRAWN TO ILLUSTRATE DIFFERENCE IN VARIABILITY IN TWO DISTRIBUTIONS, WHOSE MEANS ARE IDENTICAL

The method of relationship. The third method of treating frequency distributions is the method of relationship. For example, we need to know how one type of ability is related to another type of ability; how one type of activity "corresponds" to, or "correlates" with, another; what effect one type of learning has on another type; how ability in mathematics is related to ability in languages, etc., etc.

Throughout the discussion of the use of averages and measures of variability we should have in mind the fact that for an adequate comprehension of the status of a group of numerical data one needs to study and interpret the *whole* distribution. Averages, and variability measures, and measures of relationship are but means of representing central tendency by statements of the most probable value of the measure in question. For example, the arithmetic mean (the commonly used "average") may be said to be the "*most probable value of a series of measures.*" The value of the correlation coefficient, "*r*" (.17, .33, .42, or what-not) may be said to be only a statement of *the most probable value of the degree of relationship* which exists between the two traits in question. If it were said that the coefficient of correlation, "*r*" for the relationship that exists between scholastic ability in mathematics and that in languages, is .70, we should be able to use this value of .70 as *a statement of the probability that as "ability in language"* increases, *so does "ability in mathematics" tend to increase.* That is, that pupils high in mathematics *tend* to be high in languages. It is desired to emphasize this precaution against the wholesale acceptance of statistical devices, and to point out the need for a thorough tabulation of the original data in such complete fashion that detailed study and interpretation may be made of the raw material. With this brief preliminary statement we shall turn at once to the treatment of the problem of "averages."

2. *Discussion of averages used in educational research*

Averages describe frequency distributions by pointing out central tendencies. The attempt to describe a classification of educational data by a single number must be an arbitrary process. The student thrown on his own resources and forced to invent a way of typifying a distribution would doubtless hit upon one of the several accepted ways of doing that. Suppose that he had plotted a frequency polygon from his data, as in Diagram 19, and had raised the questions — What is the most evident "central tendency" of these data? What is their most characteristic or typical feature? It is evident that the most outstanding characteristic is shown by the high point or peak existing in the frequency polygon. Since the height of the vertical ordinate in each case represents the number of measures, this high point means that a larger number of pupils solved eight problems correctly than any other particular number of problems. The corresponding point on the scale, then, may be called: —

I. THE MODE

First method of pointing out central tendencies. The mode is simply that value on the scale which occurs most frequently. It is clear that we have here a rough device for indicating the typical tendency of a mass of data. The value that occurs the most frequently obviously points out *central tendencies*, provided a large enough number of measures is included in the frequency distribution to make it a representative or "random" sample of the total group. We shall clear up the question of "sampling" in a later chapter, but may point out now that the number of measures in a distribution is large enough to form a "random sample" when it reaches such a number that the addition of another similar group of measures will not cause a fluctuation in the magnitude of the "average" that is computed from it. Under such a condition, the mode roughly typifies the distribution in question.

The student should recognize two distinct problems arising in connection with the use of the mode in interpreting his data. (1) It may be used only as an approximate "inspection" average. Inspection of the frequency polygon reveals the modal value (or modal values if there should prove to be more than one distinct peak in the polygon). This specific modal value, which is the mid-value of the class-interval that contains the largest number of measures, depends — within a certain range over the scale — upon the *size* and *position* of class-intervals. To a considerable extent, with distributions of limited numbers of cases, and with distributions decidedly unsymmetrical in shape, this crude inspection mode is *an unstable average.* On the whole we should caution against using it for any purpose except as a very rough aid in the preliminary inspection of the frequency distribution, as an aid in "characterizing the type," — picking out summit points and central tendencies in the frequency curve.

(2) The term "mode" should be technically reserved for the "theoretical mode" (introduced by Professor Karl Pearson in 1902), which is thoroughly mathematical in its origin. It was noted above that the mid-value of the class-interval containing the largest frequency depends upon the selection of the size and position of class-intervals. It should be emphasized that the larger part of our statistical work in school research is done on a distinctly limited number of measures, and with unrefined measuring instruments. If we will postulate the increase of the number of measures to a number relatively large, and an increasing refinement of the measurement itself, then the frequency polygon (or column diagram), which we draw to represent our actual recorded data, may be said to approach continually a "continuous" or "ideal" frequency curve as a limit. That is, the smooth frequency curve (for example, Diagram 23, from the data of Table 11) represents the ideal situation, — the law that would be obtained by refined measurement of a very large number of cases.

Practically, however, we cannot attain to sufficiently refined measurement of an infinitely large number of cases, and we have to content ourselves with a theoretical fitting

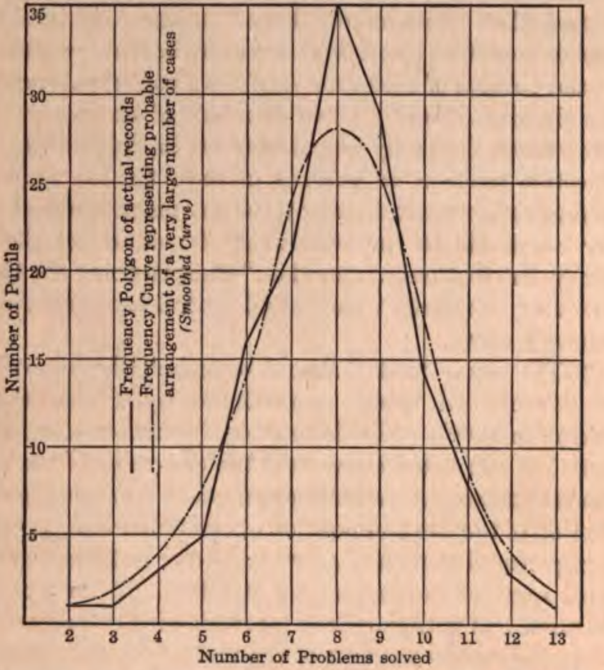


Diagram 23. Comparison of Plot of Actual Scores of 137 Pupils in Solution of Algebra Problems with "Smoothed Curve," representing a Probable Arrangement of a Very Large Number of Cases

of some frequency curve, whose equation is known, to the actual measurements. This takes the student at once to the advanced theory of curve-fitting, the thorough understanding of which implies a considerable amount of mathematical training. Thus, the discussion of the calculation of the "true mode" is clearly beyond the scope of the present work.[1]

Pearson's empirical rule for calculating mode. Fortunately, most of our distributions in educational research are but "moderately skewed," — that is, the measures are largely concentrated somewhere near the central portion of the range. For such distributions (for example Diagram 23, from data on Table 12) Pearson has given us an empirical rule for quickly calculating an approximation to this mode, which will very closely approach the true mode. It depends, however, on the previous computation of the arithmetic mean and the median. (These averages will be taken up next in this discussion.) This may be expressed as: —

$$\text{The mode} = \text{Mean} - 3\ (\text{mean} - \text{median}). \qquad -6 \quad 2-4$$

+ ?

That is, with moderately unsymmetrical distributions the median, mean, and mode stand in such a relation that the median is always about one third of the distance from the mean towards the mode. Applying this to our illustration in Diagram 20 we find, mean = 72.6; median = 75.64; difference between them = 3.04. Therefore the approximate mode is 81.72.

To give an estimate of the closeness with which the mode calculated by the use of this empirical relation approaches the "true" mode we give on page 104 two tables from Yule.

$72.6 - 3(-3.04) = 72.6 + 9.12 =$

81.7

II. THE MEDIAN

Second method of pointing out central tendencies. If the commonest measure or value on the scale is the most evident

---

[1] Complete directions are given in the complete bibliography in the Appendix concerning methods of finding the mathematical literature covering the theory of curve fitting.

TABLE 12. COMPARISON OF THE APPROXIMATE AND TRUE MODES IN THE CASE OF FIVE DISTRIBUTIONS OF PAUPERISM (PERCENTAGES OF THE POPULATION IN RECEIPT OF RELIEF) IN THE UNIONS OF ENGLAND AND WALES *

Year	*Mean*	*Median*	*Approximate mode*	*True mode*
1850..........	6.508	6.261	5.767	5.815
1860..........	5.195	5.000	4.610	4.657
1870..........	5.451	5.380	5.238	5.038
1881..........	3.676	3.523	3.217	3.240
1891..........	3.289	3.195	3.007	2.987

* Yule, *Jour. Roy. Stat. Soc.*, vol. LIX, p. 122. (1896.)

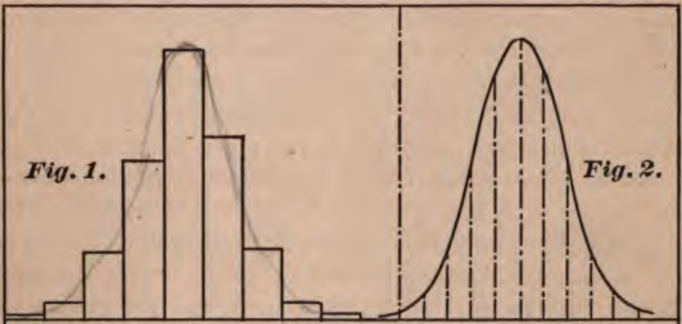
TABLE 13. COMPARISON OF THE APPROXIMATE AND TRUE MODES IN THE CASE OF FIVE DISTRIBUTIONS OF THE HEIGHT OF THE BAROMETER FOR DAILY OBSERVATIONS AT THE STATIONS NAMED †

Station	*Mean*	*Median*	*Approximate mode*	*True mode*
Southampton..	29.981	30.000	30.038	30.039
Londonderry...	29.891	29.915	29.963	29.960
Carmarthen...	29.952	29.974	30.018	30.013
Glasgow.......	29.886	29.906	29.946	29.967
Dundee.......	29.870	29.890	29.930	29.951

† Distributions given by Karl Pearson and Alice Lee, *Phil. Trans.*, A, vol. CXC, p. 423. (1897.)

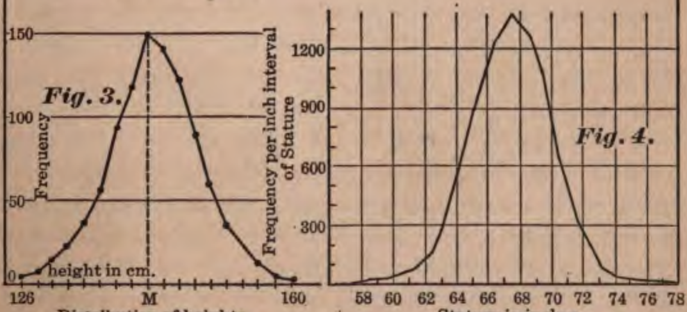
method of pointing out central tendencies in a distribution, the second is plain: find some pertinent *middle* value. Such a value is *the median*, defined rigorously as (*that point on the scale of the frequency distribution*, on each side of which one half of the measures falls.) It will be helpful to the student to do his thinking strictly in terms of the linear scale which represents the frequency distribution. The completeness with which the student refers to a scale all of his work with frequency distributions will be determined largely by the number of cases involved, and the distribution of their respective values. For example, we face two distinct problems in averaging.

Continuous series of measures. First — we have to do with two distinctly different kinds of measures in educational research: continuous series of measures, and discontinuous series of measures. A continuous series of measures is one in



Distribution of Intelligence Quotients of 905 unselected children, 5 to 14 years of age
(After Terman, p. 66)

"Normal" Frequency Distribution



Distribution of heights of 12 year old Boys
(After Whipple, p. 80)

Stature in inches
Frequency distribution of Stature for 8585 Adult Males born in the British Isles
(After Yule, p. 89)

Diagram 24. Comparison of Form of Distribution of Human Traits with "Normal Probability" Curve which the quantities are subject to any degree of division. For example, the arithmetical ability of a class of boys, as shown by the scores made on tests or by their class marks; their heights, weights and other anthropometrical measurements; in fact nearly all anthropometrical and social attributes such as we meet in educational research. We shall comment in detail in a later chapter on the form of the distribution of such human traits, but we may point out here, in order to illustrate the point of the discussion, that most human measurements have been found to conform *roughly* to some such smooth curve as is given in Diagram 24, Fig. 2.

The base line of this curve represents, in each case, the status of the trait in question, for a *very large number* of persons. The method of testing such ability results in integral scores, it is true, but in each case *these integral scores represent the mid-values of various class-intervals on the scale.*

For example, Tables 14 and 15 give the scores obtained by two groups of eleven pupils in a test for ability in factoring.

Each of these scores, 24, 23, 22, etc., means that the pupil had solved 24, or 23, or 22 problems, and was working on the next. That is, we let the integral score 22, for example, *represent* a distance in the scale, say the distance (or class-interval) from 21.5 to 22.5, or from 22.0 to 22.99. We spoke of it, in the previous chapter, as the *mid-value* of the class-interval. Thus we see that such measures form *continuous* series, — that if we refine our methods of testing we will get scores of 22.1, 22.2, etc., instead of 22, 23, 24.

Discontinuous series of measures. On the other hand, although most of our measurements are of the foregoing type, we do meet discontinuous series in our study of educational problems. For example, all our records of attendance contain gaps, — whether by classes, schools, or grades; the salary schedules of teachers contain distinct gaps, — we advance

SCORES OBTAINED BY TWO GROUPS OF ELEVEN PUPILS IN A TEST FOR ABILITY IN FACTORING

TABLE 14. GROUP I

Number of problems right	*Scale*	*Number of pupils*
24		1
23		1
22		1
21		1
20		1
19		1
18		1
17		1
16		1
15		1
14		1
Total	..	11
13	..	1
Total	..	12

With 12 measures, median = 18.5
With 11 " " = 19.

TABLE 15. GROUP II

Number of problems right	*Scale*	*Number of pupils*
24		1
23		1
22		1
21		
20		
19		
18		1
17		
16		1
15		
14		
13		
12		
11		
10		1
9		1
8		1
7		1
6		1
5		1
Total	..	11

teachers by jumps of $25.00, or $50.00, or $100.00, etc. It should be clear, however, that *for purposes of pointing out central tendencies* these measures may best be distributed along a scale and grouped, each group thus representing a distance on the "salary scale."

The second definite problem that has to be clear to the student who wishes to grasp sound methods of "averaging" takes account, first, of the differences in proper methods to use in the case of small numbers of measures, as opposed to large numbers, and, second, of the shape of the frequency distribution. This latter point takes account of the degree to which the measures are concentrated at different points on the scale, — whether near the middle or at the extreme ends. The two points must, however, be discussed together.

general

With small numbers of measures (perhaps 10 to 20 or 30), and anything but a very symmetrical distribution over a fairly short range, the wisdom of using *any* average to typify the measures is questionable. Rather than do this, the *whole* distribution should be presented and discussed in detail. Furthermore, the form of the distribution or the way in which the measures are concentrated at particular points on the scale may render any single measure decidedly fictitious. For example, suppose that the distribution showed a large proportion of measures largely concentrated at the very end of the range, but with decided numbers scattered throughout the entire range. The attempt to find some *one* typical measure to point out the central tendency of these measures must result in a partially fictitious statement of affairs. On the other hand, with the distribution of algebra scores shown in Diagram 19, a middle value, say 8, in Table 11, typifies the group very well. So, in turning to the discussion of the finding of the median, a central-most value, we should take up the discussion with a full recognition of the limitations of such single measures in typifying distributions of certain kinds.

We said that the mode was an "inspectional average." In the same way, the *median* is a *counting average*. Its determination includes two steps: (1) the arrangement of the measures in serial or rank order, placing the largest one first and the smallest one last or vice-versa; (2) the counting in of the measures from one end to determine the point on each side of which half of the measures fall. The specific computation of the median depends upon whether the measures are arranged in a simple series or in a grouped frequency distribution.

Computation of the median

(A) With the measures in a simple series. By a simple series we mean a distribution of values on a scale, each of which values occurs once. Thus Tables 14 and 15 give simple series of an odd number of measures, 11. In Table 14 it is clear that the middle-most measure, the sixth (19) is the median of the series, regardless of whether we define the median carefully as the point on the scale on each side of which there is an equal number of cases, or as the middle measure. Many people have been defining the median as the middle-most measure in the series. Obviously if we add a twelfth measure (say one case of 13 problems in Table 14) we now have no middle measure. We are forced to assume that the median is the value half-way between 18 and 19, or 18.5. This latter way of defining the median assumes that the median is the $\frac{(N + 1)\text{th}}{2}$ measure in the series (*e.g.*, $\frac{11 + 1}{2} = 6\text{th measure}$). We shall define it throughout this discussion as a *point on the scale on each side of which N/2 measures* are found to fall.

Table 14 therefore offers no very real difficulty in typifying the distribution, — *the measures are uniformly dis-*

tributed by units of one. In Table 15, however, the 11 measures are scattered over a wider range. Now the middle measure is 10, the median under the $\frac{N+1}{2}$ definition. Adding a measure, say 21, makes our total 12, with no middle measure but a hypothetical median at 13, half way between 10 and 16. It should, of course, be stressed that with 11 measures, any "average" is a questionable measure of central tendency.

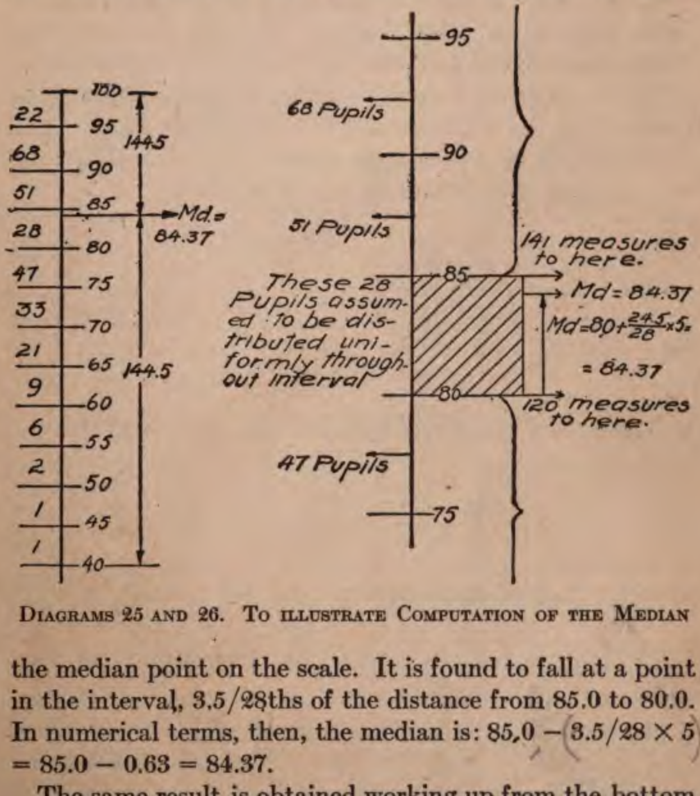
(B) With the measures grouped in a frequency distribution. As the number of measures becomes larger (30 or 40, perhaps, and upward) we are forced to group our measures in a frequency-distribution. For purposes of computing the median we now make an important assumption: *the measures in any class-interval are distributed uniformly throughout the interval, but may be represented by the value of the mid-point.* The computation may now be illustrated by the distribution in Table 16.

TABLE 16. DISTRIBUTION OF MARKS IN LATIN GIVEN TO 289 HIGH-SCHOOL PUPILS

Class-interval	*No. of pupils*
95.0–100.00	22
90.0– 94.99	68
85.0– 89.99	51
80.0– 84.99	28
75.0– 79.99	47
70.0– 74.99	38
65.0– 69.99	21
60.0– 64.99	9
55.0– 59.99	6
50.0– 54.99	2
45.0– 49.99	1
40.0– 44.99	1
	N = 289

Half the measures, *i.e.*, $N/2 = 144.5$. Therefore we wish a point on the scale on each side of which there are 144.5

measures. Counting down from the top the three-class-intervals 95.0–100.0, 90.0–94.99, and 85.0–89.99 contain 141 measures. That is, 141 measures have values greater than 85.0. In the class-interval 80.0–84.99, there are 28 measures, assumed to be distributed *uniformly* throughout the interval. Diagrams 25 and 26 show graphically the method of finding



DIAGRAMS 25 AND 26. TO ILLUSTRATE COMPUTATION OF THE MEDIAN

the median point on the scale. It is found to fall at a point in the interval, 3.5/28ths of the distance from 85.0 to 80.0. In numerical terms, then, the median is: 85.0 − 3.5/28 × 5 = 85.0 − 0.63 = 84.37.

The same result is obtained working up from the bottom of the scale. Thus, class-intervals, 40.0–44.99 to 75.0–79.99 inclusive (or from 40.0 to 80.0) contain 120 measures. We wish the point on the scale on each side of which there are 144.5 cases. Therefore we need to go up into the class-interval 80.0–84.99, 24.5/28ths of the entire distance in the intervals. In units on the scale this means 24.5/28 × 5 added to 80.0, which is the value of the lower limit of the scale, = 84.37 as before.

It will be noted that, to define the median as the point on the scale on each side of which there are $N/2$ measures, makes it possible to compute the median from either end of the scale and secure a constant value. This calls attention to the fact that the definition of the median as the $(N + 1/2)$th measure leads to inconsistent results. For example, in the computation of the following simple problem: —

$\frac{N+1}{2}$th (and some thing)

Class-interval	*Frequency f*	
20.0–24.99	17	40
15.0–19.99	23	
10.0–14.99	29	
5.0– 9.99	21	26
0.0– 4.99	5	
Total	95	

47.5

2 95.52 47.52

Working from the 20–25 class-interval downward the median equals: —

$$15.0 - \frac{48-40}{29} \times 5 = 15 - 1.379 = 13.621$$

Working upwards from the 0–5 class, the median is: —

$$10.0 + \frac{48-26}{29} \times 5 = 10 + 3.793 = 13.793$$

Thus, computing the median from one end of the distribution gives 13.621; from the other, 13.793. The method used should give the same result, regardless of the direction of computation. It is suggested here that the student should always check his work by counting in from both ends of the distribution. Using the method of computing the median, adopted here, the work checks up as follows:

(*a*) Working from 20–25: median equals the point on the scale on each side of which there are $N/2$ or 47.5 measures. Therefore

$$Md = 15 - \frac{47.5 - 40}{29} \times 5 = 15 - 1.30 = 13.70$$

(*b*) Working upward from the bottom of the distribution: —

$$Md = 10 + \frac{47.5 - 26}{29} \times 5 = 13.70$$

Summary of steps in computing median. In concluding the discussion of the median let us summarize the steps in its computation for the frequency distribution.

First: compute $N/2$ measures.

Second: Beginning at either end of the distribution, say the lower end, count the number of measures included in all class-intervals *to the interval that contains the median.*

Third: From $\frac{N}{2}$ measures subtract the total number below the interval (obtained in step 2). This number of measures is the *number* that is needed to be included from the next interval to bring the computation *to* the median point on the scale.

Fourth: Divide this remainder by the number of measures in this interval (containing the median). This is the *proportion* of the total measures in the interval that are needed to bring the computation to the median point.

Fifth: Multiply this ratio by the number of *units* in a class-interval. The product is the number of units on the scale that need to be added to the value of the lower limit of the class-interval to give the median.

Sixth: Add this number to the value of the lower limit of the class-interval. This is the median point on the scale. Ease of computation and checking will be facilitated by expressing the value of the lower and upper limits of class-intervals as whole numbers, 80.0, 85.0, 90.0, etc., instead of 79.99, 84.99, 89.99, etc.

This whole process can be duplicated from the upper end of the scale by subtracting instead of by adding.

III. The Arithmetic Mean

Third method of pointing out central tendencies. We designated our first method of pointing out central tendencies, the mode, as a rough "inspectional average"; our second method, the median, as a "counting average." It is clear that either of these methods take account but indirectly of the VALUES of the measures in the distribution. That is, the mode is determined by the *number* of measures that happen to be concentrated most largely at a certain point, — that is, the mode is a "position" average. Similarly, the median takes account of the actual VALUES of the measures only in their serial or rank order arrangement. The remaining steps in the computation of the median recognize each measure equally with all other measures. For example, in Table 10, and Diagram 20, the extreme measures of the distributions have equal weight with all other intermediate measures in the middle part of the range.

We have definite need, however, for a measure of central tendency which will take account not only of the position of each of the measures, but also of their actual numerical value. Such a measure is the arithmetic mean, which is very generally called the "average," or "arithmetic average." It should be pointed out here that the term "average" should be regarded as a class term which *will include all* of the various measures of central tendency that we are discussing in this chapter, and not as applying specifically to any one of them.

Definition and computation of the arithmetic mean

The arithmetic mean may be defined as the *sum of the values* of all the measures in the distribution, *divided by the number* of measures. Throughout this book we shall let M represent the arithmetic mean of the distribution, m represent the *value* of any measure, and N the number of cases. Thus the formula for the arithmetic mean becomes

$$M=\frac{\Sigma m}{N}$$

We must next make clear the distinctions which arise in connection with the problem of averaging by the arithmetic mean, — namely, the computation of the *simple* and *weighted* arithmetic means, considered in connection with the distribution of measures, *first*, in the simple (ungrouped) series, and *second*, in the grouped frequency distribution.

1. **The computation of the arithmetic mean with the measures reported at their true values; i.e., in ungrouped or simple series.** This may be illustrated by introducing the following table: —

TABLE 17. ANNUAL COST PER PUPIL FOR INSTRUCTION IN ENGLISH IN 10 CITIES, WITH LONG METHOD OF COMPUTING THE ARITHMETIC MEAN OF THE SIMPLE SERIES

City	*Cost*	*Frequency* f	*Frequency × measure* $f \times m$
A	$46	1	46
B	42	1	42
C	57	1	57
D	71	1	71
E	51	1	51
F	61	1	61
G	50	1	50
H	22	1	22
I	31	1	31
J	21	1	21
		10	10)452(45.2 =simple arithmetic mean

It will be noted that the data of Table 17, although forming a simple series (each measure occurring at its actual value) also represent a *simple frequency distribution*, the frequency of each value being one. The arithmetic mean that is computed from such a series is *a simple arithmetic mean.*

Table 18 also presents an illustration of the simple series, — that is, no grouping of measures has been done, and each measure appears at its true value. In this case, however, we are dealing with a WEIGHTED frequency distribution, because the frequency of occurrence of measures of any particular value is, in many cases, greater than one, — 46 at 3 cents, 26 at 4 cents, etc. In this weighted frequency

TABLE 18. COST PER PUPIL-RECITATION OF TEACHING ENGLISH IN 148 KANSAS CITIES *

Cost per student-recitation in cents. "The Measure" m	*Number of school systems. "Frequency" f*	*The measure × corresponding frequency f.m.*
12	1	12
11	1	11
10	1	10
9	5	45
8	3	24
7	6	42
6	9	54
5	23	115
4	26	104
3	46	138
2	26	52
1	1	1
Total $N =$	148	$148\overline{)608}$ 4.11 = the *true* weighted arithmetic mean of the above distribution

---

* One case marked "above 12" omitted in all computations on these data. Monroe, W. S., *Cost of Instruction in 148 Kansas High Schools.* Bulletin no. 2, Bureau of Educational Measurements and Standards, Kansas State Normal School, Emporia, Kansas.

distribution there has been no approximation, however, for each "class" in the distribution is a single unit, — 46 cities actually paid 3 cents a pupil-recitation, 26 cities 4 cents, 23 cities 5 cents, etc.

The weighted arithmetic mean. The mean that is computed here is *called* a *weighted* arithmetic mean, because certain values occur more frequently than others. In this case, however, the student should note that it is a *true mean*, just as the simple mean computed from the simple frequency distribution in Table 17 is a true mean. Furthermore, theoretically there is no difference in the principle underlying the computation of the simple mean and the weighted mean. In both, the value of each measure is multiplied by the frequency of occurrence of that measure, the products are added, and the sum is divided by the number of measures. The expression for the weighted arithmetic mean now becomes:

$$M = \frac{\Sigma fm}{N}$$

where m represents the numerical value of any measure, f the corresponding frequency of occurrence and N the total number of measures. In the actual computation of the simple mean we merely add the values of the separate measures, and divide by the number of measures. In Table 17 each of the measures has been reported as having a frequency of 1, merely to make clear that there is no *theoretical* difference between the simple and weighted mean, and that, with the former as well as with the latter case, in which the data consist of *ungrouped* measures, the computation results in a *true* mean. The computation of the weighted mean by multiplying the actual value of each measure by the corresponding frequency involves a large amount of numerical labor. It will be shown later that this labor may be very materially cut down by a short method of computation, when dealing with either the simple or the weighted arithmetic mean.

2. The computation of the arithmetic mean with the measures grouped in the frequency distribution. In the previous discussions we have noted that in grouping measures in class-intervals we make two fundamental assumptions, — (1) that the measures are distributed *uniformly* throughout the interval, and (2) that for computation purposes they are all numerically *represented by* the value of the mid-point of the class-interval. For the measures in Table 18, the effect of this is illustrated by the following grouping of the measures: —

TABLE 19. THE 148 MEASURES OF TABLE 18, GROUPED IN CLASS-INTERVALS OF 2 UNITS EACH

Cost per student-recitation in cents. "The measures" The class-interval	*Mid-point of the interval*	*Number of cities "Frequency" f*	*The measures × their corresponding frequencies f·m*
1– 2	1.5	27	40.5
3– 4	3.5	72	252.0
5– 6	5.5	32	176.0
7– 8	7.5	9	67.5
9–10	9.5	6	57.0
11–12	11.5	2	23.0
		148	148)616.0 4.16 =approximate weighted arithmetic mean of the distribution

We note that grouping the classes of the original distribution in this fashion changes the "average" cost per pupil recitation by five cents, a difference from the true mean of about one per cent. With a more symmetrical or "skewed" distribution we would have found that grouping even two consecutive intervals together would have affected the mean more considerably. With distributions that are fairly symmetrical, however, we see that the "grouping" of class-intervals changes the mean but slightly. The decision as to grouping of data, size of class-interval, etc., must depend on the data in hand. In each, there is no need of further grouping. In the case of data like those given in Table 21, showing the distribution of the percentile efficiency of 365 students in a test for visual imagery, it would be a waste of time to compute the mean of the entire ungrouped distribution. Since we have a range of 100 per cent, it will be convenient to divide the distribution into 20 class-intervals of five per cent each. The frequency distribution is then as given herewith in Table 21.

Table 20 illustrates the effect of grouping on the size of the arithmetic mean and median, by giving results for the true mean computed from the 123 original measures of Table 9, p. 83, ungrouped, and the approximate mean for grouping these same measures in class-intervals of three units, five units, and ten units respectively. These results may be tabulated as follows:

TABLE 20. EFFECT OF GROUPING ON THE SIZE OF THE ARITHMETIC MEAN OR MEDIAN

True mean or median (i.e., data in original units of 1 each)	*Mean or median with data grouped in class-intervals of*		
	3 units	*5 units*	*10 units*
Arithmetic:			
Mean..................72.17	72.6	73.11	72.89
Median.....................	75.64	76.05	75.10

The frequency distributions and polygons representing these data, as given in Table 9, and Diagrams 20 and 21, are seen to be but moderately skewed. For such a type of distribution it is clear that grouping the measures changes the "average," either mean or median, relatively little.

Table 21 gives the detailed computation of the arithmetic mean of the achievement of 365 college students in tests for visual imagery, by the traditional or LONG method. This method groups all measures in a class-interval at the mid-point. This example makes it clear that the computation of a mean by this method is unwieldy. Short methods of computation are therefore desirable.

TABLE 21. EFFICIENCY OF 365 COLLEGE STUDENTS IN TESTS FOR VISUAL IMAGERY

(*The long method of computing the weighted arithmetic mean*)

Class-intervals	*Frequency f*	*Value of mid-point m*	*fm*
95.0–100.0	8	97.5	780.0
90.0– 94.99	2	92.5	185.0
85.0– 89.99	9	87.5	787.5
80.0– 84.99	8	82.5	660.0
75.0– 79.99	24	77.5	1860.0
70.0– 74.99	16	72.5	1160.0
65.0– 69.99	33	67.5	2227.5
60.0– 64.99	11	62.5	687.5
55.0– 59.99	35	57.5	2012.5
50.0– 54.99	18	52.5	945.0
45.0– 49.99	59	47.5	2802.5
40.0– 44.99	20	42.5	850.0
35.0– 39.99	56	37.5	2100.0
30.0– 34.99	20	32.5	650.0
25.0– 29.99	18	27.5	495.0
20.0– 24.99	6	22.5	135.0
15.0– 19.99	12	17.5	210.0
10.0– 14.99	4	12.5	50.0
5.0– 9.99	4	7.5	30.0
0.0– 4.99	2	2.5	5.0
	365		365) 18,632.5
			51.05 = weighted arithmetic mean

3. A short method of computing the arithmetic mean. The short method to be presented to the student shortens the labor of multiplication by making three conditions: (1) that we treat the class-interval as a unit of 1 on the scale, instead of as an aggregation of many units; (2) that we assume the value of any measure or class-interval as the estimated mean, or as the one which contains the estimated mean; (3) that we compute the difference between the true mean and the estimated mean, rather than compute the true mean itself. Let us illustrate it first by application to the simple series of cost data reported in Table 17.

TABLE 22. MEAN FOR TABLE 17, RECALCULATED BY SHORT METHOD

Cost	*Estimated mean*	*Deviation from estimated mean in actual units*		
		+	*0*	−
51		5		
42				− 4
57		11		
71		25		
46	46		0	
61		15		
50		4		
22				−24
31				−15
21				−25
	Total = 10 measures	60		−68 60 $\Sigma d = -8$ $\frac{\Sigma d}{n} = -.8$

$M = \text{Estimated mean} + \frac{\Sigma d}{n}$

$= 46 + (-.8)$

$= 45.2$, as by the long method

180
10

[illegible]

[illegible]

The practicableness of the use of the short method in saving time in a simple series is doubtful. This simple illustration is included here to make clear the principle underlying the use of the method in the case of the frequency distribution. With grouped series it has practical value as a labor-saving device. Let the student note clearly, however, before turning to the more complicated illustration given below, that the short method merely estimates the mean, and *then adds a correction* (c) *which is the arithmetic mean of the deviations from this estimated mean.* $c=\frac{\Sigma fd}{N}$ and the formula for the mean becomes: —

$$M = \text{estimated mean} + \text{correction}$$

$$M = \text{estimated mean} + \frac{\Sigma fd}{N} \times \text{number of units in the interval}$$

in which d is the deviation of any class-interval from the interval containing the estimated mean. Obviously the method will hold true regardless of the frequency of occurrence of the measures. We pointed out in the previous sections that there is no theoretical difference between the simple and weighted mean. It will be noted that the multiplication can now be done mentally. In Table 23 we apply the method to the data of Table 21.

The value of the whole method lies in the fact that it is a time- and labor-saving device. We estimate the "assumed mean," and compute *mentally* the correction that has to be made to the assumed mean.

The example in Table 23 illustrates the use of the short method. The entire distribution of 365 cases is first grouped in 20 class-intervals, each interval having a range of five per cent. This step results in the frequencies 8, 2, 9, 8, 24, etc. These are then totaled, giving 365. Instead of next multiplying the mid-point of each interval (a three-place

TABLE 23. EFFICIENCY OF 365 COLLEGE STUDENTS IN VISUAL IMAGERY

Class-intervals	*Frequency f*	*Deviation from the assumed mean interval d*	*Frequency × deviation fd*	
95.0–100.0	8	10	80	
90.0– 94.99	2	9	18	
85.0– 89.99	9	8	72	
80.0– 84.99	8	7	56	
75.0– 79.99	24	6	144	
70.0– 74.99	16	5	80	
65.0– 69.99	33	4	132	
60.0– 64.99	11	3	33	
55.0– 59.99	35	2	70	
50.0– 54.99	18 164	1	18	
45.0– 49.99	59	0	703	703
40.0– 44.99	20 142	−1	− 20	−444
35.0– 39.99	56	−2	−112	259
30.0– 34.99	20	−3	− 60	
25.0– 29.99	18	−4	− 72	
20.0– 24.99	6	−5	− 30	
15.0– 19.99	12	−6	− 72	
10.0– 14.99	4	−7	− 28	
5.0– 9.99	4	−8	− 32	
0.0– 4.99	2	−9	− 18	
	365		444	

259 divided by 365 = .71; .71 × 5 = 3.55, the correction to be added to assumed mean to get the true mean. The true mean = the assumed mean + correction.

Assumed mean = 47.50
Correction = 3.55
True mean = 51.05

number) by the corresponding frequency, we estimate the class-interval, 45.0 to 49.99, the mid-point of which most closely approximates the position of the true mean. This can be determined by inspecting the frequency distribution, counting up from one end until half the cases are included. Practice in this "scanning" of the distribution will give skill in closely approximating the true mean.

The labor involved in mental multiplication will be further reduced by taking the assumed mean at the portion of the distribution at which the measures are most heavily concentrated.

Use of class-intervals with the short method. The next step consists of tabulating the number of units distant that the mid-point of each class-interval is from the mid-point of the interval containing the assumed mean, 47.5. These distances are called deviations "*d*," and in the example are: interval 50.0–54.99 is 1 unit above, or larger than, 45.0–49.99, therefore its deviation is +1; *d* for 55.0–59.99 is +2; for 40.0–44.99 is −1; for 35.0–39.99 is −2, etc. Thus, the whole short method merely treats the class-intervals as units of 1 instead of 5 or whatever they may be.

With the traditional method of finding the weighted mean we would next multiply the mid-point of each class-interval by its frequency. Instead, we now *multiply the "deviation"* of each class-interval from the assumed mean by the frequency of the class, 8 × 10, 2 × 9, 9 × 8, etc., giving the column headed *fd*. Two points must now be kept in mind, — first, the *fd's* occupy the same place in the computation by the short method that the *fm's* do in the traditional method; second, that *having assumed a mean, all of the deviations above the mean will be positive, and all below the mean will be negative.* If we were dealing with the *true mean* of the distribution, the *sum of the positive deviations should be equal to the sum of the negative deviations.* Since only in rare cases does the true mean fall at the mid-point, it will always be *necessary to ADD to the estimated mean* a "correction" (denoted "*c*"). Just as we find the *sum* of the *fm's* in the long method, so do we find the *ALGEBRAIC SUM* of the *fd's.* That is, we total the positive deviations and the negative deviations, and subtract the smaller from the larger. This difference is then the *total amount* that the mid-points of all the class-intervals deviate from the *assumed mean*. However, we wish the *average amount* of the deviations of the mid-points of the class-intervals from the assumed mean. This will be the correction "c'." The average amount is found by taking the arithmetic mean of the sum of the deviations, — in the example given this amounts to dividing the difference of the positive and negative deviations, 259, by the total number of cases, $N = 365$. This gives a correction $c' = +.71$, which means that the assumed mean is smaller than the true mean by .71 of the range of the class-interval. Note that to find the true mean we do *NOT* add *this value* of the correction to the assumed mean, for this value has been computed on the basis of the class-interval of 1, instead of 5. Therefore the true correction is $.71 \times 5 = 3.55$, which, if added to the assumed mean, 47.5, will give the true mean, 51.05. The accuracy with which the mean worked by this method checks the mean worked by the longer method, depends merely upon the number of decimal places to which the arithmetic work is carried by the two methods. In Table 5 the problem discussed above is worked by the long method to permit a comparison of the two methods.

Summary of steps in the computation of the arithmetic mean by the short method. In conclusion, let us summarize the method as follows: —

1. Group the original measures in a frequency distribution.
2. Total the frequencies.
3. Estimate the interval that contains the mean. The value of the mid-point of this interval is the value of the estimated mean.
4. Treating each class-interval *as a unit*, record the number of units that the mid-point of each class-interval *deviates* from the estimated mean, indicating as positive all intervals whose mid-value is greater and as negative all those whose mid-value is smaller than the estimated mean. These distances will nearly always be less than 10, because most distributions

will contain less than 20 intervals, and the estimated mean will be taken approximately in the middle of the distribution. A safe rule is to take the estimated mean in the heavily concentrated portion of the distribution. In this way the mental multiplication will involve smaller numbers.

5. Multiply each deviation (d) by its corresponding frequency (f) *taking account of signs.*
6. Find the algebraic sum of the positive and negative deviations.
7. Divide this sum by (N) the number of measures. This gives the correction (c'), which is the arithmetic mean of the deviations from the estimated mean, in units of class-intervals.
8. Multiply c' by the number of units in an interval, giving c.
9. Add c to the estimated mean to get the true mean.

For most of the "averaging" problems of school research the three methods discussed in the foregoing pages suffice. A detailed analysis will be given later of the specific use of various methods of averaging. There are two problems involving the computation of averages, when time rates or rates of increase are in question, that have to be treated by special averaging methods. We shall turn to these next.

IV. The Harmonic Mean

The averaging of time rates. At the present time the measuring movement in education consists largely in the establishment of "norms of attainment" in the various school subjects, and for various levels of scholastic development. We have many "grade norms" in handwriting, spelling, reading, and arithmetic in the elementary school, and for algebra in the secondary school. The "norm" is taken to be the "average" performance (expressed as so many words written in one minute, or read in one minute, etc.) of large groups of pupils found at the different years of school life who are actively taking work in the various studies. "Average" performance has quite universally been taken to be the arithmetic mean of the performances of the individual pupils. Generally this has been true irrespective of the conditions of work involved in the testing. In fact, it seems quite clear that there is no general recognition of the fact that there is *an issue involved in the averaging of time rates.*[1]

We shall therefore call attention to certain points in the use of statistical averages which may have been overlooked by workers in educational research. It is desired to establish the following points: —

1. That there are two distinctly different methods of averaging time rates:
   (*a*) Averaging by the arithmetic mean of the rates;
   (*b*) Averaging by the harmonic mean of the rates;
2. That with given material average performances computed by the two methods will not be comparable;
3. That these two methods imply two different units of computation, "the unit of work" and the "unit of time";
4. That a method of averaging must be selected appropriate to the unit of computation which is being used, — with the unit of work we must use the harmonic mean of the rates (the arithmetic mean of the absolute times); with the unit of time we must use the arithmetic means of rates.

To get the problem clearly before us let us use the following simple illustration: —

Suppose a group of five boys to have been tested for speed of solving the algebra problems used in the writer's Test 1, Series A, by assigning a definite amount of time (2 minutes) and noting the amount of work done. Let us express the results in two ways : (1) express the efficiency as "the number of problems worked correctly in one minute"; (2) express the efficiency as the "number of seconds required to solve one problem" (assuming the problems to be uniform in difficulty, on which basis the test was designed). The

---

[1] This issue was first pointed out to the writer by Dr. L. P. Ayres. The writer later met the problem in his own work and is alone responsible for the present method of treatment.

first method expresses performance as a rate, or in terms of a unit of time; the second method expresses performance in terms of a unit of work (the amount of time required to do a unit of work). Let us now find the "average" performance of the results of the testing, by computing the arithmetic mean (the method commonly used) of the individual records in the two series. The computation is as follows: —

Number of problems solved per minute	*Number of seconds required to solve one problem*
12	5
10	6
8	7.5
6	10
4	15
5)40(8 problems solved on the average per minute 8)60(7.5 sec. required to solve one problem	5)43.5(8.7 sec. required to solve one problem, or an average rate of 6.897 problems per minute

Formula of the harmonic mean. The question arises, why is not the time required to solve one problem as obtained by one arithmetic mean the same as the time required when obtained by the arithmetic mean in the other series? It is noted that the rate as determined by the two methods differs as much as fifteen per cent. The answer to the question is: *The two series are not comparable until reduced to the same base.* The base required is: What part of a minute is required to solve one problem? In the second series this is the base used (*i.e.*, the number of seconds required to solve one problem). Each member of the first series needs to be reduced to that base. In other words, the reciprocal of each measure should be obtained instead of the rates themselves, and these should be averaged by the arithmetic mean. *This amounts to finding the harmonic mean of the series of rates.* We may define the harmonic mean as follows: *it is the reciprocal of the arithmetic mean of the reciprocals of the in-*

dividual measures of the series. It may be expressed by the following formula: —

$$\frac{1}{H}=\frac{1}{N}\Sigma\left(\frac{1}{m}\right)$$

where N = number of cases, and m represents any individual measure. It should be stressed that the harmonic mean of the rates is the same thing as the arithmetic mean of the corresponding time. The work now checks up as follows: —

Number of problems solved per minute	*Reciprocal of number of problems solved per minute*	*Number of seconds required to solve one problem*
12	.08333	5
10	.10000	6
8	.12500	7.5
6	.16667	10
4	.25000	15
5)40(8	5).72500(.1450	5)43.5(8.7 sec.
8)60(7.5 sec. required to solve one problem, *according to the arithmetic mean of the rates*	$\frac{1}{.1450}$ = 6.897 problems can be solved in one minute = rate. 6.897)60(8.7 sec. required to solve one problem *according to the harmonic mean of the rates*	required to solve one problem. Rate = 6.897 problems solved per minute, *according to the arithmetic means of the absolute times*

It has been recognized that the harmonic mean of a series of rates will always be less than the arithmetic mean. This simple problem shows that it will be less by as much as fifteen per cent *with distributions of large variability.* Naturally the two means approach each other in value as the variability decreases.

It is clear that there are two distinctly different ways of approaching the problem of establishing standards of attainment in various mental or physical abilities. They are plainly to be distinguished on a basis of the unit involved, the unit of work, or the unit of time. To repeat them here, they are: (1) the unit of work: How much time is required to do a unit of work? (2) the unit of time: How much work is done in a unit of time? It will be agreed that in order to get comparable average measures we must use the same method of averaging individual records in the two series.

Proper method of averaging with each of the different units. Granted that there are two distinctly different methods of averaging time rates (*i.e.*, two different units of computation), and that results computed by the arithmetic mean on the basis of one unit are not comparable with those computed on the other unit, the question arises: Which method of averaging should be used; (1) with the unit of work; (2) with the unit of time? It must be recognized at the start that the taking of an average to represent or typify large numbers of measures is in a sense an arbitrary process. It is merely an attempt to select one numerical index (out of several possible ones) which shall represent adequately the status of the entire group. To state our problem clearly let us turn to the stock problem of the men rowing a boat at different rates. We may then adapt the conclusion of the matter to our own problem of educational measurement.

First, the unit of work: Assume A and B each are to row one mile (or work one problem) and the time is to be taken. A rows the mile in 7.5 minutes, *i.e.*, he rows the mile at the *rate* of 8 miles an hour. B rows the mile in 5 minutes, *i.e.*, he rows 1 mile at the *rate* of 12 miles an hour. Together they row 2 miles in 12.5 minutes, or 1 mile in 6.25 minutes, or *at the average rate of 9.6 miles an hour*. On the other hand if we took the arithmetic mean of the two rates themselves, 8, 12, we would conclude that the *average rate* of rowing was 10 miles an hour. This would assume that the actual elapsed time for the two men over each mile was 12 minutes, or 6 minutes for the average of the two. This is incorrect, for the actual elapsed time over each mile for the two men was 12.5 minutes or 6.25 minutes for the average of the two. We may sum up the statement of the procedure in this way: with a unit of work used as a basis the average rate must be such that the two men will row 2 miles (or solve 2 problems, if we wish to substitute the algebra test in place of the rowing problem) every 12.5 minutes. In terms of averages this means that *with any problem stated in terms of units of work, we must average "rates" by the harmonic mean.* That is, in order to give consistent results the rate per minute, or per hour, etc., must be turned into "elapsed time required to do a unit of work" and the corresponding arithmetic mean computed. In other words we must satisfy the equation, —

$$r_{ave.} \times t_{ave.} = T$$

or, the average rate multiplied by the average time required to do a unit of work equals the total elapsed time. As illustrated above in the problem, to average a series of measures by the harmonic mean: (1) take the reciprocal of each measure; (2) find the arithmetic mean of their reciprocals; and (3) find the reciprocal of this arithmetic mean. This is the average rate as computed by the harmonic mean.

Second, the unit of time: Assume A and B are each to row *one hour* (or work algebra problems one hour). A actually rows 8 miles in one hour, *i.e.*, he rows at the *rate* of 8 miles an hour. B actually rows 12 miles in one hour; *i.e.*, he rows at the *rate* of 12 miles in one hour. Thus in one hour they both row 20 miles, and their *average rate* on this basis is the arithmetic mean of 8 and 12, or 10 miles. This rate we contrast with 9.6 miles an hour, as computed by the harmonic means of the same rates 8 and 12.

From the above illustrative problem we find that we cannot compute average rates by the arithmetic means of the individual rates, irrespective of the unit of computation.

We find also clear illustration of the fact that, although the taking of an average measure is only a makeshift as a representative or type for the individual measures or a series, — in other words that the selection of an average is, in a sense, an arbitrary process, — yet each average has a particular function and can be applied only in connection with particular units of computation.

V. THE GEOMETRIC MEAN

This latter point can be made still more evident by reference to the specific use of the *geometric mean*. We may define this mean as the *nth* root of the product of the separate measures in the series, — that is,

$$M_G = \sqrt[n]{(x_1 x_2 x_3 \;.\;.\;.\;.\; x_n)}.$$

There has been practically no use made of the geometric mean in educational research, in spite of the fact that with problems of averaging rates of increase the average to use is the geometric mean. For example: —

Suppose an individual's performance, as shown by testing, had improved fifty per cent in ten practice periods, say in ten weeks. What is the average weekly rate of improvement? Is it five per cent, as shown by the quotient of the total improvement divided by the number of weeks? On the contrary it is found by taking the 10th root of 1.50 and subtracting the initial efficiency, *i.e.*, $\sqrt[10]{1.5} - 1$, which gives us 4.1 per cent. In other words a weekly improvement of 4.1 per cent will increase the efficiency by 50 per cent in ten weeks. It is clear that we cannot take the arithmetic mean of such geometrical increase as the above. To do so in this case would give a total improvement of 63 per cent, instead of 50 per cent.

The geometric mean, practically adapted only to the solution of short series, can easily be computed by the aid of logarithms. Thus the logarithm of the geometric mean of a series of measures is the arithmetic mean of the logarithms. The expression would read thus: —

$$\log M_G = \frac{\Sigma \log x}{N}.$$

Steps in the computation of the geometric mean. The steps in the computation of a geometric mean are therefore as follows: —

1. Find the logarithm of each of the measures.
2. Find the arithmetic mean of the series of logarithms.
3. Find the number corresponding to the arithmetic mean of the logarithms; — this is the geometric mean of the original series of measures.

Another illustration of the use of the geometric mean is given below: —

1. Suppose a group of boys to have gained skill in practicing shooting, 90 per cent in 3 months. What has been the average gain each month? Not $\frac{90}{3} = 30$ per cent but

$$\sqrt[3]{1.90} - 1.0 = 1.\dot{2}4 - 1 = 24 \text{ per cent.}$$

That is, at the end of the first month their gain in efficiency is 24 per cent + 100 per cent = 124 per cent of their initial efficiency, which was 100 per cent. At the end of the second month they have gained 24 per cent of 124 per cent = 29.76 per cent, which added to 124 per cent gives an efficiency of 153.76 per cent. The third month they gain 24 per cent of 153.76 per cent = 36.24 per cent, and their final efficiency is 190 per cent of their initial efficiency, a gain of 90 per cent as stated above.

Thus we have described and discussed the computation of five specific averages, which are available for use by students of educational research: —

1. The mode (approximate or true), a "position" average.
2. The median, — a "counting" average.
3. The arithmetic mean, — an arithmetic average based on the value of each measure.
4. The harmonic mean, — an average for use in averaging time rates.
5. The geometric mean, — an average with particular uses in averaging rates of increase.

Of these, by far the greater use is made of the median and arithmetic mean. It is necessary next to establish the proper function, limitations, and specific use of each of these methods of averaging.

VI. Statistical Fallacies and Mooted Points in Averaging

As students of education have turned to the use of quantitative methods, many fallacies have been evident in their manipulation of statistical devices. These are found in connection with methods of averaging, of measuring dispersion, of measuring correlation, and of determining the reliability of measures. We shall discuss each of these pitfalls in the use of statistical methods as we take up each phase of the work. We must point out here, then, certain typical fallacies in averaging, and proceed to the thorough analysis of the proper use of averages. Fallacies of averaging have been of two kinds: (1) those in which the *wrong average* has been used (*e.g.*, the arithmetic mean instead of the harmonic mean); (2) those in which an incorrect use has been made of an average, — that particular average being the proper one to use in the given problem (*e.g.*, the use of the simple instead of the weighted mean).

A. Use of wrong average. In this type of mistake in averaging we find: —

1. *Averaging time rates by the arithmetic mean.* We have

already shown that rates of achievement, obtained by the "unit of work" method of testing, may not properly be averaged by finding the arithmetic mean of the rates themselves, but rather by finding the arithmetic mean of the corresponding "times" (*i.e.*, by the harmonic means of the rates).

2. Averaging rates of increase by the arithmetic mean. In the last section we pointed out the difficulty of defining the average of a series of percentage increases by taking their arithmetic mean. We found rather that the *nth* root of the product of the measures (*i.e.*, their geometric mean) might better be taken to represent the average status of the measures.

B. Incorrect use of an average. Of this second general class of incorrect uses we find the use of the simple arithmetic mean for the weighted arithmetic mean. To this may be added the error of assuming that the measures of a distribution are distributed *uniformly* throughout the range of the distribution. The most evident of such mistakes found recently is: —

1. Taking the arithmetic mean of the extremes of a distribution as the "average" of the measures in the distribution. This is one of the most patent of fallacies in averaging. It is illustrated clearly by the following data. A State commission appointed to survey the State's higher educational institutions, collected data on the occupancy of classrooms in three state institutions by this method, — viz: "The maximum occupancy of any room (the maximum number of students regularly in the room at any period of the week) plus the minimum occupancy, divided by 2, equals the average occupancy." Furthermore, to obtain the occupancy ratio for any building or group of buildings, they obtained the "average" for each room and took the simple arithmetic mean of these averages. This report gives no complete data upon which to compare the actual occupancy with these fictitious figures. They are obviously fictitious, however, and are based on two unsound assumptions: first that the distribution of classroom occupancies is uniform, and second, that the frequency of occurrence of each size of classroom is constant. A recent report calling attention to this fallacy says: "Room 32 of a similar building in another state had a minimum occupancy of 1, and a maximum of 56." The actual occupancy was then stated as follows: —

Class	*Size of class*	*Number of times room is used by each class f*	*f·m*
A	56	3	168
B	23	2	46
C	19	3	57
D	6	6	36
E	5	4	20
F	4	2	8
G	1	3	3
Total	114	23	338

Average occupancy = 23)338(14.7

By the method of taking the arithmetic mean of the extremes the occupancy is, $\frac{56 + 1}{2} = 28.5$. In brief, such an error in the use of averages is really caused by using the simple mean instead of the weighted mean, in that it assumes that the frequency of use of each classroom is the same. It also mistakenly assumes that the sizes of class are distributed uniformly over the entire range. This obviously is not true in school practice.

Which mean to use. The question as to which arithmetic mean to use (simple or weighted) in averaging educational data is one of great importance at the present time. The answer can be given the student only in terms of the relation between the nature of the data at hand and the purpose of interpretation. *Use that method of averaging which will give the truest picture of the central tendencies evident in your data.* In the foregoing example the actual occupancy of classrooms is clearly typified better by the weighted mean than by the simple mean of the two extreme measures of the distribution. Educational conclusions based on such a method as the latter must necessarily hamper the progress of scientific education. Let us give some concrete illustrations of the use of the simple and weighted mean.

The most frequent demand for "averages" is in connection with the attempt to measure various aspects of school efficiency. Our figures are stated, for example, in terms of the achievement of pupils determined by testing; unit (average) costs of various school activities; average age, experience, training or salary of teachers; average amount of time devoted to this, that or the other subject of study, etc. Measurement of classes, schools, and systems of schools gives distributions of data that are to be expressed in terms of central tendency. Shall we express this by weighting every class, school, or system with the number of pupils in each, number of teachers, number of rooms, etc., or by taking the *simple* arithmetic mean of the records of classes, schools, buildings, or systems? This amounts to asking in the case of the achievement of pupils, — *what is the basic unit in our data* — the pupil or the class? the ability of the pupil regardless of training, or the specific type of training to which he has been subjected?

Take the case of testing pupils' efficiency in algebra, as the writer has done it in 50 school systems. The number of algebra pupils per school varied from 30 to 100; the average achievement varied among schools by very large amounts on any one test. Shall the score of the school of 100 pupils be weighted 100, and the score of the school of 30 pupils, 30? Or, shall they each be regarded as of equal weight with all the others, and the simple mean be computed? The answer must be made in terms of the basic unit — the unit clearly is the *class*, not the pupil. We are testing the *result* of the pupil's training in algebra, his skill in doing a specific thing he has been trained to do. We are testing the results of *a score of types of training*, and these are the basic units. Contrast this situation with the determination of average height of school boys, the average age of teachers, etc. Here the basic unit is very clearly the individual boy or teacher, not the class into which he or she may be grouped, and the records of classes, schools, groups, etc., should be weighted by the number of individuals.

Another commonly occurring problem nowadays is the school-cost problem. We meet a series of heating costs computed say, for 20 school systems, by buildings, in units of, "per cubic foot," "per class room," or "per pupil in average daily attendance." In such a problem we should first classify buildings in groups in terms of like heating conditions, — similar heating apparatus, like number of rooms, etc. If this is impossible then unit heating costs for city systems clearly should be computed with the basic unit taken to be the classroom, cubical contents, or number of pupils, and not the building.

Homogeneity of data. Still another very important problem of averaging is raised in connection with the question of "homogeneity of data." It is fundamental to the sound treatment of numerical data that we *include in any one statistical group only individuals who have been subject to the same conditioning factors*. For example, the attempt to compute the "average" salary of all teachers in a school system cannot possibly result in a clear statement of "average" salary which will definitely be comparable to that computed for another system. The "average" in this case is computed from a distinctly non-homogeneous group of persons, — elementary teachers, secondary teachers, elementary principals, secondary principals, supervisors of grades and special subjects, assistant superintendents, superintendent, and other special administrative officers. To secure comparable measures we clearly must average separately for each statistical group, making sure that each is made up of persons whose salary status is determined by the same set of causes. The student should guard constantly against the fallacy of computing averages from non-homogeneous data. He will meet series of data, continually, in which he has included items that are caused by conditions qualitatively different, and which should be eliminated from the group.

For example, suppose that we have tested classes of pupils in arithmetic. In a class of 20 there are three who do not attempt any problems of the test. Should we sum the scores of the class, and divide by 20 or by 17? The arithmetic mean will be distinctly different in the two cases, and our interpretation of comparisons correspondingly so. Such cases must be decided by reference to the question of "homogeneity of data." If the class is under our immediate control it will be possible to tell if these three pupils in *intellectual capacity, previous training and physical condition on the day of the test are qualitatively different* from the other 17 members of the class, who solved problems varying in number from 3 to 18. Comparison of the scores in various tests in the same subject will also help us to decide. If they prove to be so, they should be eliminated from the group and the average computed from the records of the 17.

Another illustration from the field of school costs will make the point clearer. A recent study on the relationship between the cost of instruction and the number of pupils taught by one teacher gives the data reproduced in Diagram 47, Chapter IX. It will be noted that the table includes all groups of data on the number of pupils taught by a teacher, from 25 to above 170. Careful examination of the table will show, however, that the investigator has two distinct groups of conditions included in his study. It is evident that for the range from 25 pupils to 80 pupils taught by one teacher there is a very high degree of relationship, *i.e.*, that as the number of pupils increases the cost decreases in a definite way. This relationship may be expressed by a coefficient of correlation of —.84. From 80 pupils throughout the rest of the table it is evident that, as the number of pupils increases there is no decrease or increase in cost, and the coefficient is practically 0. The investigator has thrown the two distinctly different groups together, and computed relationship for a non-homogeneous group. His coefficient of —.47 and his averages and measures of variability are largely fictitious for that reason, and conclusions based on them are of questionable value.

Averaging "samples." Before leaving this introductory discussion of the uses of particular averages we should refer briefly to the effect of averaging inadequate "samples" of our total mass of data. In educational research we are constantly forced to form conclusions from a relatively small group of data. How large, for example, should a group be, or how many times should a test be given to permit *general conclusions* to be drawn concerning similar individuals in the mass or similar testing work? In other words, how many cases must we have to give us an "average," typical of a very large number of similar individuals? To illustrate the point: suppose that we wish to determine the spelling ability of 20,000 pupils in a city system, representing the achievement in part by some measure of "average" attainment. It is not expedient to test all of them. How many pupils shall we test to get a "random sample"? A common sense way to define such a sample is this: A sample of any total population is "random" when numerical coefficients, for example averages, computed from any number of samples similarly selected and of similar size will be approximately constant. (The more technical phases of "sampling" in statistics will be discussed later.)

Functions and limitations of particular averages. We have thus introduced the subject of the functions and limitations of averages by a concrete exposition of particular difficulties that the student will meet in pointing out central tendencies in his data. It should be recalled here that these difficulties are of two types: (1) those which may involve the taking of the distinctly wrong average; (2) those which involve the application of any average to non-homogeneous data, or to an inadequate sample, or to an improper determination of the basic unit. The second point has been discussed completely enough to lead to a thorough presentation of the former point. Therefore, we turn next to the question of the properties of each of the five averages, their proper functions, their limitations, and the specific purpose for which each should be used.

Enough has been said to make it clear that the process of averaging is one of selecting the best single quantity to characterize the central tendency of a distribution; that any average that is used must have certain properties which will show it to be a good representation of type.

Summary of essential properties of a valid average.[1] It will aid the discussion to list here the essential characteristics of representative averages: —

1. If it is to be completely representative of the entire distribution, it must be contributed to by all the measures of the distribution.

---

[1] The writer has been aided in making a complete summary of these properties by Yule's discussion, *Introduction to the Theory of Statistics*, chapter on averages.

2. It should be purely quantitative, — defined by the numerical data alone, and should not involve the judgment of the observer.
3. It should be so constructed as to be relatively simple in computation.
4. It should be stable; that is, it should be of such a nature that representative samples taken from the total population will give a fairly constant average value. All other factors being equal, that average which gives the smallest fluctuation in value as we take different samples from the total group, is the best average to use.
5. An average must not be much displaced by slight changes in the arrangement of the frequency distribution. References to the discussion of the arrangement of data in the frequency distribution will make clear the importance of getting an average that will be fairly stable, regardless of the size or position of class-interval that is selected. Furthermore, the average must be as little as possible affected by errors in observation.
6. Since the purpose of averaging is to point out clearly central tendencies to the reader, the average which is selected should be of such simple and definite nature that the lay reader will grasp easily its typifying significance. In this characteristic the geometric and harmonic means show themselves to be poor averages, the arithmetic, median, and mode being much more easily understood. Complete success in using an average must depend on the student and the reader being able to think clearly in terms of the average.
7. From the standpoint of mathematical treatment, in the refined use of averages, it is important that an average be susceptible of algebraic manipulation. For example, it has been repeatedly pointed out that it should be possible to express an average obtained from the combination of two or more samples of the same data in terms of the averages of each of the samples.

VII. USE OF THE DIFFERENT MEASURES OF CENTRAL TENDENCY

It will now be possible to come to some agreement concerning the proper use of averages by checking each against the foregoing list of essential properties.

Function of the mode as a measure of type. Taking up our properties in order, these conclusions seem evident:

1. The mode is not contributed to by all the measures. On the contrary it may be determined by a relatively small proportion of the total number of measures, concentrated in one class.
2. It is quantitative in the sense that it is defined by the frequency of the largest class.
3. The empirical mode is an inspection average, and thus is the easiest of all the averages to determine. Furthermore, it may be determined without any detailed knowledge of the extremes of the distribution except that the frequency of measures there is small. On the other hand, the theoretical mode is the most difficult to compute of any of the averages, depending on the most advanced theory of "curve-fitting."
4. It is more unstable than the arithmetic mean or median in its fluctuations, due to the taking of different samples from a given group of data.
5. In any but closely symmetrical distributions it is relatively unstable in the way in which it depends very closely on the method of grouping of the class-intervals. The manner in which it fluctuates is illustrated by Diagrams 20 and 21, as we change the size and position of the class-interval. For fairly refined work it is evident that the mode is too unstable for effective use.
6. The mode has the advantage of being the most easily comprehended of any of the averages. It is the "newspaper average"; the average of the man on the street, and for the lay reader has a clearer meaning than most of the other averages. Here it finds its principal function in describing skewed distributions of many class-intervals, with distinct concentration of measures in certain class-intervals. Furthermore, it serves a good purpose in the graphic representation of measures, being marked by distinct peaks in the frequency polygon.
7. It is clear that the empirical mode (the only one in which the student of education is interested) has no mathematical significance, and is not susceptible of algebraic treatment as is the arithmetic mean.

In résumé, it should be clear that the empirical mode is only a rough inspection average; that it may be indicated to the reader as one means of pointing out central tendencies; but that its capacity for representing the central tendency is very limited. Dependence on it beyond preliminary inspection of a distribution is not to be recommended, except in very symmetrical distributions.

The geometric mean as a measure of central tendency. With the exception of problems involving the averaging of rates of increase, the student of educational research will have comparatively little need for using the geometric mean. Its computation is rather laborious; it is not readily comprehended by the lay reader (not having come into popular use); and its mathematical properties are abstract, although valuable in certain forms of problem work.

The principal function of the geometric mean is found in treating data which involve rates of increase, and which thus take the form of geometric series. For example, in problems in averaging increases in population, attendance in school, growth in the teaching staff, budget, etc., average status can be more consistently defined by means of the geometric mean.

A second valuable property of the geometric mean is found in connection with the discussion of index numbers or ratios. It may be said that the mathematical properties of the geometric mean establish the superiority of that mean over that of the arithmetic mean or the median, in averaging such index-numbers.

Use of the harmonic mean in measuring central tendency. In connection with the discussion of the harmonic mean on pages 126–131, its specific function as an average of progress rates was pointed out. This valuable property of the harmonic mean should be kept in mind, and brought into use in all problems of that nature.

The median as a measure of central tendency. With reference to the median, the following conclusions seem evident:

1. The median is contributed to by all the measures of the series, the magnitude of each, however, being taken account of only indirectly. That is, the median is an average depending on the *serial order* of values and on the actual numerical value only as it determines this serial arrangement.
2. It is quantitative, being defined at least indirectly by the values of the measures.
3. It has the great advantage that it is the most easily computed of all the numerical averages; it is a "counting average," depending for its determination on (*a*) the serial arrangement of the measures (with the use of the frequency distribution this is a necessary step of the computation of the arithmetic mean also); (*b*) the counting in of half the measures to reach the median point on the scale.
4. Fluctuations in the size of the median may be larger with the taking of small samples. At the same time the median may give a more stable average from small samples, due to fluctuation in the size of extreme values. In this particular it should be pointed out that the median is affected less by the extremes of the distribution — that is by unusually large or small measures — than is the arithmetic mean, which takes full account of these values. The student must decide carefully, in connection with his specific distribution, whether the "average" should or should not be contributed to by unusually large or small values. If they are regarded as important the arithmetic mean is the best representative of central tendency; if not, then the median is the better measure of type. Again, the location of the median depends only partially on a small group of measures; in this, it differs distinctly from the mode. However if it happens that the measures in a distribution are largely concentrated in a few intervals, it may result that the median (falling at a point on the scale at which many measures are concentrated) will be very indefinite.
5. With the types of distribution commonly met in educational problems, the median is but little subject to fluctuation with rearrangement in the size and position of class-intervals. Reference to Diagrams 20 and 21 shows the relatively stable position of the median in the distribution of fairly large numbers, with a form not more than moderately skewed.
6. The median must rank high in the ease with which its mean-

ing may be grasped by the lay reader. Partly for this reason, it is being adopted rapidly by students of education.

7. The median does not lend itself to algebraic treatment. (*a*) The median of component parts of a distribution cannot be expressed in terms of the median of whole distribution; this is true because the distribution depends on the *form* of the component distributions, and not on their medians alone. (*b*) No theorems can be expressed for the median values of measurements subject to error.

Use of the arithmetic mean as a measure of central tendency. Applying the criteria of the essentials of a valid average to the arithmetic mean we find that it outranks all the others as a sound measure of central tendency.

It conforms to all the stated properties for a desirable mean as listed above. It is definitely and numerically defined; is based on all the measures; is popularly known and commonly used, hence will always be readily grasped by the lay reader; is very easily calculated (in this it ranks high as a mean, *e.g.*, the short method of computing the mean is also a necessary step in the determination of the standard deviation and of the correlation coefficient); the aggregate and the number of cases are sufficient to enable the computation of the mean, *i.e.*, the specific individuals do not need to be treated; and in adaptation to algebraical treatment it has a great advantage over the other means. For example, important properties of this mean are: —

1. The algebraic sum of the deviations from the arithmetic mean equals 0;
2. The average of a series may readily be expressed in terms of the means of component parts of the series. From this it can be deduced that the approximate value of a mean in a frequency distribution is the same whether we assume that all the values in any class are identical with the mid-value of the class-interval, or that the mean of all the values in the class is identical with the mid-value of the class-interval;
3. The mean of all the sums and differences of corresponding measures in the two series (of equal number of measures) is equal to the sum or difference of the means of the two series.

The arithmetic mean is also characterized by the fact that the sum of the squares of the deviations of measures from the mean is a minimum. The arithmetic mean has properties of fundamental importance in the field of mathematical statistics, especially in connection with the theory of errors and the theory of probability (*e.g., the arithmetic mean can be shown to be the most probable value of a series of measures*). Accidental errors of observation tend to neutralize each other around the arithmetic mean. The error of the average is considerably smaller than the error of a single measure, and the accuracy of the arithmetic mean varies directly with the square root of the number of the measures. The median and mode have no similar properties.

ILLUSTRATIVE PROBLEMS *

1. Find the arithmetic mean and the median of each of the following distributions. In the computation of the mean use the short method.

1. ACHIEVEMENT OF 5TH GRADE PUPILS IN SPELLING 25 WORDS FROM COLUMN L OF AYRES'S "SCALE FOR MEASURING SPELLING ABILITY"

No. words spelled correctly	*Frequency* (*f*)
25	7
24	5
23	11
22	14
21	21
20	13
19	8
18	7
17	5
16	3
15	5
14	3
13	2
N =	104

2. DISTRIBUTION OF MONTHLY SALARY PAID TO TEACHERS OF ENGLISH IN 149 KANSAS HIGH SCHOOLS

Monthly salary	(*f*)
$120.0–124.99	3
115.0–119.99	
110.0–114.99	1
105.0–109.99	1
100.0–104.99	2
95.0– 99.99	
90.0– 94.99	3
85.0– 89.99	10
80.0– 84.99	30
75.0– 79.99	36
70.0– 74.99	31
65.0– 69.99	20
60.0– 64.99	8
55.0– 59.99	1
50.0– 54.99	1
N =	

---

* These illustrative problems are quoted from Rugg, H. O., *Illustrative Problems in Educational Statistics*, published by the author to accompany this text. (University of Chicago, 1917.)

3. Achievement of Pupils in solving Problems of Test 3, "Standardized Tests in 1st Year Algebra"

No. problems solved correctly	*f*
21	3
20	5
19	3
18	11
17	16
16	21
15	29
14	20
13	17
12	10
11	5
10	3
9	7
8	3
7	2
N =	

4. Distribution of Monthly Salary paid to Teachers of Science in 147 Kansas High Schools *

Monthly salary	(*f*)
$135.0–140.00	1
130.0–134.99	3
125.0–129.99	4
120.0–124.99	4
115.0–119.99	2
110.0–114.99	10
105.0–109.99	7
100.0–104.99	26
95.0– 99.99	8
90.0– 94.99	16
85.0– 89.99	22
80.0– 84.99	15
75.0– 79.99	15
70.0– 74.99	5
65.0– 69.99	4
60.0– 64.99	2
N =	

* Data from Monroe, W. S., *Cost of Instruction in Kansas High Schools.*

CHAPTER VI

THE MEASUREMENT OF VARIABILITY

SECOND METHOD OF DESCRIBING A FREQUENCY DISTRIBUTION

Need for measures of variability. It has been pointed out in the last chapter that the average of a distribution cannot possibly *completely represent* the measures of the distribution. At best, it is but a partial measure of type, arbitrarily selected to represent central tendency. We have indicated with what relative degree of success the different averages do this. Frequently the student of education will have to compare two distributions in which the average is closely the same, but in which the FORM of the distribution is very different. This calls attention to the need of interpreting our data only after careful examination of *both* the entire distribution and the frequency polygon plotted from it.

For example, Diagram 22, Chapter V, is drawn to represent the achievement of two classes. The average, as shown by the arithmetic mean or the median, is the same in both distributions. If one should compare the two distributions on the basis of the average achievement alone his interpretation concerning the outcomes of teaching in the two classes would most certainly be wrong. This is evident by a study of the characteristic differences in the two frequency distributions: (1) the "RANGE" in the one case is nearly as long as in the other; (2) on the other hand, the *measures* in one distribution are very much *more concentrated* near the middle of one group than of the other. One could say, for example, that the "middle half" was distributed over a portion of the scale not much more than half as large in one case as in the other. This certainly means that the teaching has in one case served to develop a rather compact group, that is, teaching emphasis has been so distributed that differences in achievement have been largely smoothed out. In the other case the teaching has resulted in a widely scattered group, certainly calling for reclassification of pupils in connection with any further learning in that subject.

In pointing out the characteristic differences between such distributions we make clear the kind of measure that is needed with which to supplement the use of an average. We need some measure which will indicate *the degree to which the measures are concentrated around the average*, or — to express it another way — a measure which will point out concretely the degree to which the *measures vary away from the average.* That is, we need *measures of variability or dispersion.*

Variability a distance on a scale. We found that a measure of central tendency, such as an average, is always expressed as "position," — *as a point on the scale.* We now find that with symmetrical distributions, a *measure of variability* is always *expressed as that distance on the scale*, which includes a particular proportion of the measures in the distribution. Although educational distributions are not perfectly symmetrical, it will be a helpful device for pointing out the degree of concentration or lack of concentration of the measures to say: "*approximately such a proportion of measures is included between such unit distances on the scale.*" We have already emphasized the importance of the term "unit" and "scale." The student now will find that his *measure of variability is nothing but a unit distance on the scale.* Of the different unit distances that we have for measuring varia-

bility, each includes a certain proportion of the measures under the frequency curve.

Four measures of absolute variability. For example, the four measures of absolute variability that we use, and the approximate proportion of the measures included within their limits, when laid off on the scale, are:—

1. The *range:* includes all of the measures in the distribution.

2. The *quartile-deviation* or *median-deviation:* when laid off on each side of the average: includes only roughly half of the measures.

3. The *standard deviation:* when laid off on each side of the average: includes approximately the middle two thirds of the distribution.

4. The *mean deviation:* when laid off on each side of the average: includes approximately the middle half of the measures.

Unit distances with normal and skewed distributions. We should emphasize the fact here, that with distributions that are not perfectly symmetrical (*e.g.*, Diagram 27), we are able to state the proportion of measures included by these unit distances on the scale only very approximately. With a symmetrical distribution, for example, the "probability curve" shown in Diagram 28, we are able to state the proportion of measures exactly. Diagrams 27 and 28 illustrate this distinction, and, although we shall take them up more thoroughly in Chapter VIII, we may point out the important features here. On these diagrams are indicated graphically and literally the chief characteristics of these measures of variability.

It will be evident to the reader that with the perfectly symmetrical distribution, Diagram 28, any unit distance may be laid off from the average (in this case arithmetic mean, median, and mode coincide) either way and include

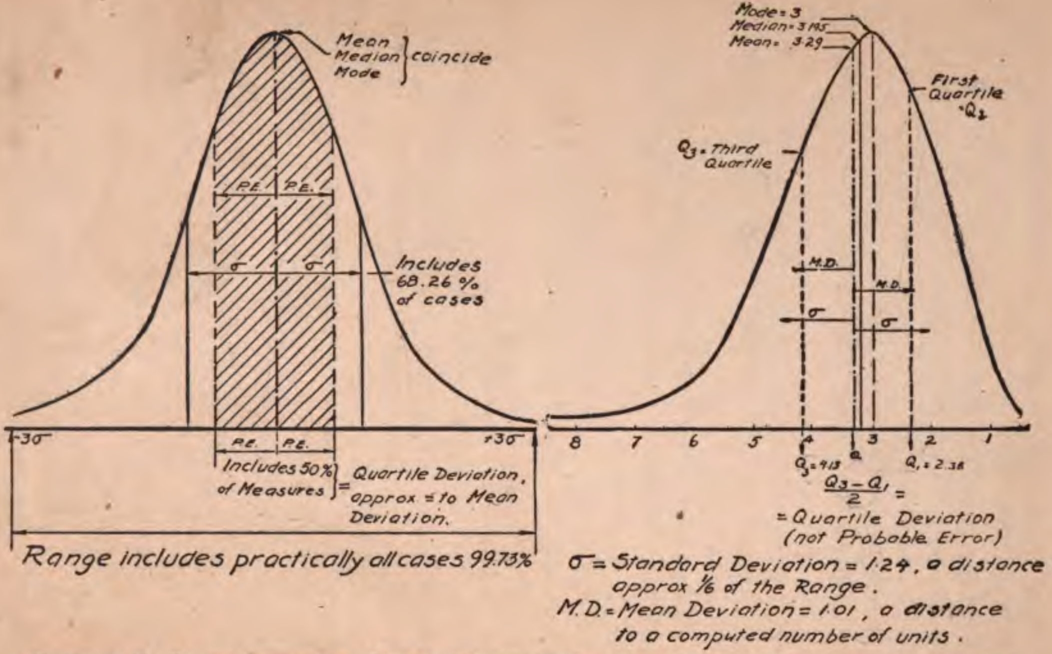


DIAGRAM 27. TO ILLUSTRATE THE USE OF "STANDARD DEVIATION," "MEAN DEVIATION," "QUARTILE DEVIATION" ON "NORMAL" AND "SKEWED" CURVES

the same proportion of cases. Thus, it will be shown in Chapter VII that between the curve, the base line, and ordinates erected at a unit distance from the mean called the *Probable Error* (*P.E.*), 25 per cent of the measures is included, or 50 per cent between *P.E.* and the curve and the base line. In this case it is clear that *on the "Normal Curve"* the *quartile deviation* (defined as half the distance between the first and third quarter points on the scale) equals the *P.E.* In the same way it will be shown that between the curve, the base line, and ordinates erected at a unit distance called *sigma* (σ), the *standard deviation*, 68.26 per cent of the measures is included. Turning to the unsymmetrical distribution, Diagram 27, we see that we cannot define our variability *rigidly* in terms of the *proportion* of measures included between ordinates erected at

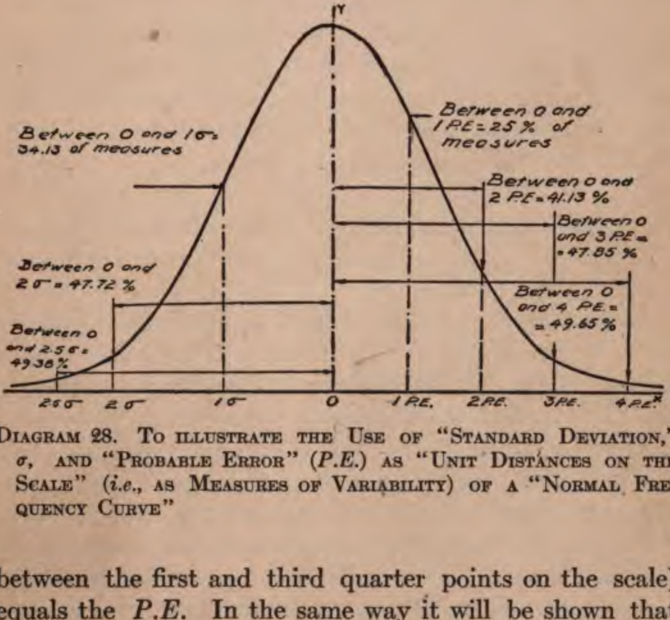


DIAGRAM 28. TO ILLUSTRATE THE USE OF "STANDARD DEVIATION," σ, AND "PROBABLE ERROR" (*P.E.*) AS "UNIT DISTANCES ON THE SCALE" (*i.e.*, AS MEASURES OF VARIABILITY) OF A "NORMAL FREQUENCY CURVE"

the *same unit distance* from the average. Nevertheless, it will be helpful to think of the variability, after it is computed, in terms of *distance on the scale*, — thus picturing to ourselves roughly the compactness of our distribution.

Leaving to Chapters VII and VIII the detailed discussion of frequency curves and their properties, we will turn to a systematic presentation of the measures of variability. We should distinguish at the start two kinds of variability: (1) *absolute variability*, typified by any one of the four measures, — the range, the quartile deviation, the mean deviation, and the standard deviation; and (2) *relative variability*, described by so-called (*a*) coefficients of variability, or (*b*) measures of skewness. The distinction will be made clear in presenting the latter two devices.

I. MEASURES OF ABSOLUTE VARIABILITY

1. The Range

An unstable measure of variability. We have defined the range as the difference, or the distance on the scale, between the largest and smallest measures. For example in the distribution plotted in Diagram 20 (classification of 3 units) the range is 19 to 97. Inspection of the diagram will show that if we eliminate one measure at the low end of the scale, the range becomes 28 to 97; if we cut off two more measures it becomes shortened to 40 to 97. This calls attention to the fact that the range is a very unstable measure of variability, in that it may depend so completely on the value of a single measure, or of a very small group of measures. Thus, the range takes no account of the *form* of the distribution — *i.e.*, the degree of concentration of measures at various points on the scale. With fairly compact symmetrical or moderately unsymmetrical distributions the investigator should always state the range, in connection with other measures of type or variability, as a rough guide to the interpretation of his data.

2. The quartile deviation, or semi-interquartile range

The middle half of the measures. It has been suggested by many investigators in this field that a convenient measure of the *form* of a distribution, *i.e.*, of the degree of concentration of the measures, would be to find how large a distance on the scale contains the middle half of the measures. Yule has called half of this distance the *semi-interquartile range*, expressed by

$$Q = \frac{Q_3 - Q_1}{2}$$

that is, *half the distance between the first and third quarter points on the scale.* Q_3 and Q_1 are thus quarter points on the scale, defined as those points above and below which one fourth or three fourths of the measures fall. Thus, the median is Q_2, the second quarter or quartile point. This calls to mind then, that the quartile points are computed for both ungrouped and grouped observations by exactly the same method as that with which we compute the median. Having computed the quartile points one might take the distance (or difference) between them as a measure of the variability. Most of our distributions are not perfectly symmetrical, and so it has become standard practice to use half this distance as the unit of absolute variability. *In reality, it is not a deviation at all*, being determined merely by counting in on the scale a given number of measures. Just as the median is a counting measure of central tendency, *i.e.*, an average, so is the quartile deviation another counting measure of central tendency, *i.e.*, a measure of variability. In computing it, however, no average is found and no particular deviation from any central point on the scale is computed. In brief, the quartile deviation is simply a convenient counting device for pointing out the *position on the scale of the middle half of the measures.*

P.E. and quartile deviation. Writers in education have often incorrectly called this measure of variability the *Probable Error*. The latter term should be reserved specifically for the treatment of that symmetrical distribution known as the probability curve. Reference to Diagram 27 shows that the probable error (*P.E.*) and the quartile deviation are the same in the probability curve. Each one is equal to such a distance on the scale that when laid off on each side of the average, it will include one half of the measures. The student should make himself thoroughly familiar with this unit of scale distance, the properties and use of which will be taken up in the next chapter. Most distributions of educational data are moderately skewed, and so it will be wise to use the term quartile deviation (Q) very generally. Yule's term, the semi-interquartile range, although having a more specific connotation, appears to be too cumbersome to obtain common usage.

Computation of the quartile deviation: (*a*) data ungrouped. The computation of Q for a short simple series is very clear from inspection of the range.

The steps may be listed as follows: —

1. Divide the number of measures by 4.
2. Count in on the distribution from either end to the point on the scale above or below which there are $\frac{1}{4}$ or $\frac{3}{4}$ of the measures. For example, in Series I, Table 24, —
   - (*a*) Q_1 is the arithmetic mean of the values of the 6th and 7th measures; $Q_1 = 71.0$; Q_3 is the arithmetic mean of the values of the 18th and 19th measures; $Q_3 = 87.5$.
   - (*b*) In Series II, $\frac{1}{4}$ of the measures is 6.5, hence the quartile points may be regarded as the 7th and 20th measures, for these are the points which theoretically have above

TABLE 24. TO ILLUSTRATE THE COMPUTATION OF QUARTILE DEVIATION. MEAN DEVIATION AND STANDARD DEVIATION FOR THE UNGROUPED SERIES.

(1) *Series I*	(2) *Series II*	(3) *Deviation from median* d'	(4) *Deviation from arithmetic mean* d	(5) fd^2
96	96	15.5	16.03	256.9
94	95	14.5	15.03	225.9
93	94	13.5	14.03	196.8
92 Q_3	93	12.5	13.03	169.8
90	92	11.5	12.03	144.7
88	90	9.5	10.03	100.6
87	89	8.5	9.03	81.5
86	88	7.5	8.03	64.5
85	87	6.5	7.03	49.4
84	86	5.5	6.03	36.4
83	84	3.5	4.03	16.2
81	82	1.5	2.03	4.1
80	81	0.5	1.03	1.1
78	80 medi-	0.5	0.03	
76	78 an=	2.5	1.97	3.9
75	77 80.5	3.5	2.97	8.8
74	76	4.5	3.97	15.8
72	75	5.5	4.97	24.7
70	73	7.5	6.97	40.6
67	72	8.5	7.97	63.5
65 Q_1	70	10.5	9.97	99.4
64	67	13.5	12.97	168.2
63	65	15.5	14.97	224.1
62	64	16.5	15.97	255.0
	63	17.5	16.97	288.0
	62	18.5	17.97	322.9
24m's	26)2079(79.97 = mean	235.0 $M.D.=9.04$	235.06 $M.D.=9.04$	26)2871.8 $\sigma^2=110.45$ $\sigma=10.51$

and below them $\frac{1}{4}$ or $\frac{3}{4}$ of the measures, that is, 6.5 or 19.5 measures.

3. Subtract Q_1 from Q_3 and divide by 2, giving Q, the deviation.

(b) For measures grouped in the frequency distribution. The steps of the computation now become exactly the same

87.5
71.

87.5
71
16.5

as those involved in the computation of the median (the second quartile point).

TABLE 25. TO ILLUSTRATE COMPUTATION OF QUARTILE DEVIATION FOR THE GROUPED SERIES

Class-interval		
95.0–100	8	
90.0– 94.99	3	
85.0– 89.99	9	=61
80.0– 84.99	4	
75.0– 79.99	24	
70.0– 74.99	13	
65.0– 69.99	26	Q_3=67.308
60.0– 64.99	12	
55.0– 59.99	27	
50.0– 54.99	13	Q=½ of this distance
45.0– 49.99	45	
40.0– 44.99	21	
35.0– 39.99	44	Q_1=37.727
30.0– 34.99	15	
25.0– 29.99	17	
20.0– 24.99	2	
15.0– 19.99	9	=51
10.0– 14.99	3	
5.0– 9.99	3	
0.0– 4.99	2	
	N=300 $N/4$=75	

1. Divide N by 4. $300/4 = 75$ measures.
2. For Q_3, there are 61 measures above 70.0. We need 75–61, or 14 measures from the 26 in class-interval 65.0–69.99.
3. Therefore $Q_3 = 70.0 - 14/26 \times 5 = 70.0 - 2.692 = 67.308$.
4. Similarly for Q_1; since there are 51 measures in the intervals 0–4.99 to 30.0–34.99 inclusive, we need 75–51, or 24 measures from the 44 in class-interval, 35.0–39.99.
5. Therefore $Q_1 = 35.0 + 24/44 \times 5 = 35.0 + 2.727 = 37.727$.
6. Therefore since $Q = \frac{Q_3 - Q_1}{2}$, we have $\frac{67.308 - 37.727}{2} =$

$$\frac{29.581}{2} = 14.791.$$

Properties of the quartile deviation. On account of the simple meaning of the quartile deviation, it is a good measure of variability to use in presenting facts for the lay reader. It further has the advantage of being the most easily computed of any of the measures of variability. In brief, Q is the inspectional or approximate measure to use in expressing variability, in the treatment of any data in which numerical precision is not necessary, or where the theory does not imply algebraic treatment. There are many opportunities to-day, in educational research, to use the quartile deviation as a measure of variability.

3. The Mean Deviation

What the mean deviation is. We pointed out that, strictly speaking, the quartile deviation is not a measure of *deviation* from a particular average. Expressed in another way this means that the quartile deviation takes but indirect account of the *form* of the frequency distribution, — of the relation between the values of particular measures and the frequency of their occurrence. There are two measures of variability that do this, however: *the mean deviation*, and *the standard deviation*. They differ only in the fact that in the former case simple deviations are averaged *without* regard to sign, and in the latter case the deviations are averaged after each has been squared, with the necessary subsequent step of extracting the square root.

The mean deviation of a series of measures is the arithmetic mean of their deviations from a selected average (median or arithmetic mean) *the deviations being summed without regard to sign.* In the computation of an average deviation, the *taking account of signs* would result in a fictitious measure of deviation, the difference between positive and negative deviations being always very small, and equal to 0 when the deviations are taken from the arithmetic mean. Therefore, to average simple deviations we are forced to disregard signs. From the *practical* standpoint, the deviations may be taken from either the arithmetic mean or from the median. The computation in columns (3) and (4), of Table 24, show that in the case of that simple series, fairly uniformly distributed as it is, the mean deviation computed from either average is the same to the second decimal place, 9.04. This will be true also with the data grouped in the symmetrical distribution, and with those not more than moderately unsymmetrical in form. From the theoretical standpoint, however, *it is proper to take the deviations from the median*, for that is the point on the scale about which the mean deviation is the least. Because it is much simpler of computation, and because of this mathematical relation, the recommendation is made that the student adopt the practice of computing deviations from the median.

Computation of the mean deviation: (*a*) data ungrouped. Let us list the steps in the computation when the data are ungrouped. Each step is illustrated by the data of Series II, Table 24: —

1. Compute the median: 80.5. ✓
2. Compute the deviation of each measure from this value, 15.5, 14.5, 13.5, etc., in column 3.
3. Sum these deviations: 235.0.
4. Find the arithmetic mean of the deviations by dividing the sum, 235, by the number of measures, 26, giving the mean deviation, 9.04. Column 4 gives corresponding deviations from the arithmetic mean, 79.97, a total of 235.06, and the same value for the mean deviation, = 9.04.

Computation of the mean deviation: (*b*) data grouped in the frequency distribution. The computation still may follow the steps given above, which may be called "the long method." The work given in Table 26 illustrates this method.

TABLE 26. DISTRIBUTION OF MARKS GIVEN TO 289 HIGH-SCHOOL PUPILS IN LATIN. TO ILLUSTRATE COMPUTATION OF MEAN DEVIATION BY LONG METHOD

Class-interval	*Mid-point of class-interval m*	*Frequency f*	*Deviation d*	*fd*
95.0–100	97.5	22	13.12	288.64
90.0– 94.99	92.5	68	8.12	552.16
85.0– 89.99	87.5	51.	3.12	159.12
80.0– 84.99	82.5	28	−1.88	52.64
75.0– 79.99	77.5	47	−6.88	323.36
70.0– 74.99	72.5	33	11.88	392.04
65.0– 69.99	67.5	21	16.88	354.48
60.0– 64.99	62.5	9	21.88	196.92
55.0– 59.99	57.5	6	26.88	161.28
50.0– 54.99	52.5	2	31.88	63.76
45.0– 49.99	47.5	1	36.88	36.88
40.0– 44.99	42.5	1	41.88	41.88
		$N=289$		2623.16
		$\frac{N}{2}=144.5$		$M.D.=\frac{2623.16}{289}$
		True median $=84.38$		$=9.08$

To make use of this long method necessitates a large amount of computation. The arithmetic labor may be cut down very materially by using the principle employed in the short method of computing the arithmetic mean. To do this here would involve these fundamental steps: —

1. Compute the total deviations about an assumed median. This can easily be done by taking the assumed median at the mid-point of the class-interval which contains the true median.

2. Correct these total deviations about this assumed median by an amount equal to the difference between the deviations about the assumed median and the total deviations about the true median.

It will be shown below that the sum of the deviations about any assumed median *must always be less than* the sum of the deviations about the true median. Hence, whatever be the relative position of the assumed and true medians, the *correction* of the deviations around the assumed median to the true median *must always be added.*

Let us contrast in Table 27, the computation by the short method with the correction applied in this way, and with the reduction to class-intervals of one unit each, but *first,* with the deviations stated in their true value, 0.62, 1.62, 2.62, etc., instead of 1, 2, 3, etc. We can then go, **second,** a step farther and compute the deviations in terms of units of 1, 2, 3, etc., and correct once for all by the short-method stated below, as in Table 28.

TABLE 27. TO ILLUSTRATE THE COMPUTATION OF MEAN DEVIATION WITH DEVIATIONS STATED IN TRUE VALUES, BUT IN UNITS OF CLASS-INTERVALS

Class-interval	f	*True-deviation of mid-points in units of class-intervals* d'	fd'
95.0–100	22	2.62	57.64
90.0– 94.99	68	1.62	110.16
85.0– 89.99	51	.62	31.62
80.0– 84.99	28	− .38	10.64
75.0– 79.99	47	−1.38	64.86
70.0– 74.99	33	−2.38	78.54
65.0– 69.99	21	−3.38	70.98
60.0– 64.99	9	−4.38	39.42
55.0– 59.99	6	−5.38	32.28
50.0– 54.99	2	−6.38	12.76
45.0– 49.99	1	−7.38	7.38
45.0– 44.99	1	−8.38	8.38
	$N=289$	$d'=$	$\Sigma d'=524.66$

True median = 84.38 $\frac{\Sigma d'}{N}=\frac{524.66}{289}=1.816=M.D.$ in units of class-intervals

$1.816\times5=9.08=M.D.$ in actual units

TABLE 28. TO ILLUSTRATE THE COMPUTATION OF MEAN DEVIATION BY SHORT METHOD. DEVIATIONS TAKEN ABOUT THE ASSUMED MEDIAN, IN UNITS OF CLASS-INTERVALS

Class-interval	*f*	*d*	*fd*
95.0–100.	22	3	66
90.0– 94.99	68	2	136
85.0– 89.99	51	1	51
80.0– 84.99	28	0	
75.0– 79.99	47	1	47
70.0– 74.99	33	2	66
65.0– 69.99	21	3	63
60.0– 64.99	9	4	36
55.0– 59.99	6	5	30
50.0– 54.99	2	6	12
45.0– 49.99	1	7	7
40.0– 44.99	1	8	8

$N=289$

True median = 84.375
Assumed median = 82.50
$c=1.88/5=.38$

$\Sigma fd=522$

Total correction = c, difference above and below true median = $.38\times 7=2.66$

Total deviations in units of class-intervals, *i.e.*, $\Sigma fd=522+2.66=524.66$

$$\therefore M.D.=\frac{\Sigma fd}{N}=\frac{524.66}{289}=1.816$$

$1.816\times 5=9.08$

Diagram 29 presents several of the class-intervals in enlarged form, together with the relative position of the true and assumed medians, and the relative sizes of the true and calculated deviations. The student is reminded again of the necessity for doing his thinking in terms of the *scale* of the frequency distribution. The diagram makes it clear that the deviation of each measure in any class-interval, when taken from the assumed median (a mid-point of a class-interval), is in error by that part of a class-interval that separates the true and assumed medians. For example, each of the 28 measures in class-interval 80.0−84.99, assumed to have a deviation of 0, actually has a deviation, from the true median (*T.M*d.), of −.38 of an interval; each of the 51 measures in interval 85.0–89.99, similarly taken at a deviation of 1, actually deviates from the true median by 0.62, etc. In other words, each of the measures above the *T.M*d. is assumed to be longer than it really is, by .38 of an interval, and each of those below the *T.M*d. is assumed to be .38 shorter than it really is. Thus, there are 141 measures

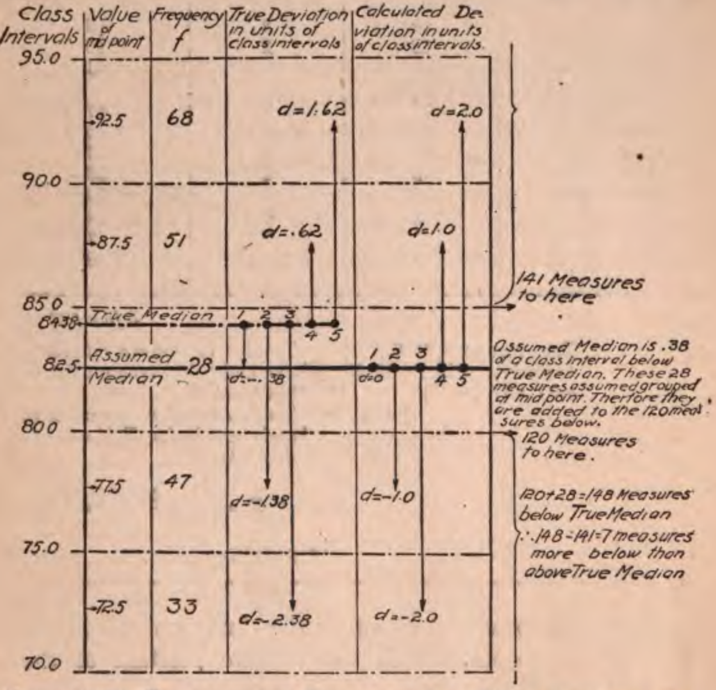


Diagram 29. To illustrate Computation of Mean Deviation by the Short Method

calculated longer than they really are, and 148 measures calculated shorter than they really are, each by .38 of a class-interval. In other words, there are 141 measures both above and below the $T.M_d.$ whose deviations are equally long or short, the effect of those above neutralizing that of those below. In addition there are 7 more measures that are short by .38 of a class-interval (7 × .38 = 2.66), making the total correction to be added to the deviations from the assumed median ($A.M_d.$) 524.66, as with the long method in Table 26.

Furthermore, there will always be more cases short than long, because the assumed median ($A.M_d.$) determines the number of cases above or below the $T.M_d.$, and the deviations taken on the side of $A.M_d.$ are always short. Thus the correction will always be added. Table 29 illustrates this point for the case in which the $T.M_d.$ falls below the $A.M_d.$

TABLE 29. DISTRIBUTION OF MARKS GIVEN TO 123 HIGH-SCHOOL PUPILS IN ENGLISH. TO ILLUSTRATE COMPUTATION OF MEAN DEVIATION BY THE SHORT METHOD, WHEN THE TRUE MEDIAN FALLS BELOW THE ASSUMED MEDIAN.

	f	*d*	*fd*	
95.0–100	1	4	4	=68 measures above the true median
90.0– 94.99	2	3	67 6	
85.0– 89.99	13	2	26	
80.0– 84.99	21	1	21	
75.0– 79.99	31	0		$A.M.=77.5$
70.0– 74.99	19	−1	−19	$T.M.=76.05$
65.0– 69.99	10	−2	−20	$c.=145/5=-.29$
60.0– 64.99	7	−3	−21	
55.0– 59.99	9	−4	−36	
50.0– 54.99	3	−5	−15	
45.0– 49.99	2	−6	−12	=55
40.0– 44.99	2	−7	−14	
35.0– 39.99	1	−8	−8	
30.0– 34.99	1	−9	−9	
25.0– 29.99	0	−10		
20.0– 24.99	1	−11	−11	
	$n=123$		$\Sigma d=222$	

The deviations of each of the 68 measures is short by .29.
The deviations of each of the 55 measures is long by .29.
The total deviations, 222, are therefore short by

$$13 \times .29 = 3.77. \qquad \frac{222 + 3.77}{123} \times 5 = M.D. = 1.836 \times 5 = 9.180$$

To express the work algebraically, let *T.M*d. equal true median; *A.M*d. equal assumed median. N_a = number of measures above *T.M*d.; N_b = number of measures below *T.M*d.; *c*, the correction = *T.M*d. − *A.M*d. no. units in interval.

$$M.D. = \frac{\Sigma fd + c(N_b - N_a)}{N}$$

In the problem, Table 27 or 28,

$$N_a = 68 \qquad T.M_{d.} = 76.05$$

$$N_b = 55 \qquad A.M_{d.} = 77.5$$

$$\Sigma fd = 222. \quad c = \frac{76.05 - 77.5}{5} = \frac{-1.45}{5} = -.29$$

$$\therefore M.D. = \frac{222 + [-.29\ (55 - 68)]}{123}$$

$$= \frac{222 + (-.29 \times -13)}{123}$$

$$= \frac{225.77}{123} = 1.836 \text{ in units of class-intervals}$$

$$\text{or,} \quad M.D. = 1.836 \times 5 = 9.180 \text{ in actual units.}$$

Steps in the computation of the mean deviation by the short method. In conclusion, we may sum up the steps in the use of the short method as follows, illustrating each step by the data of Table 28.

1. Tabulate the frequencies by class-intervals (as was done by the computation of the arithmetic mean) and total them, *i.e.*, $N = 289$.
2. By methods described in Chapter V compute the true median (84.38).
3. Select as the assumed value of the median the mid-value of the class-interval *that contains the true median* (80.0–84.99); therefore the assumed median = 82.5.

4. Find "c," the correction, or difference between the mid-value of the class-interval and the true median in units of class-intervals. This is $\frac{84.38 - 82.5}{5} = .38$.
5. In the illustrative problem this means that the sum of the deviations of the class-intervals *about the assumed median* will be in error from the sum of the deviations *about the true median* by an amount equal to the difference between the number of measures above and below the true median, *multiplied by the difference between the true and assumed medians* (*i.e.* by the correction "*c*"). Therefore, next compute the difference in the number of measures above and below the true median; 7, in this case.
6. Since each of these measures is in error by .38, the total deviations times their corresponding frequencies, when computed from the assumed median are in error by $7 \times .38$ or 2.66. Therefore next compute the amount of this total correction.
7. Tabulate the deviations (*d*) of the mid-value of each class-interval from that of the assumed median (precisely as was done in the computation of the arithmetic mean).
8. Multiply each frequency by its respective deviation, giving the column of *fd's*.
9. Find the sum of the deviations (*fd's*), *without regard to sign. Remember that this sum is taken about the* assumed median.
10. Add the total correction, from 6 above (in the problem this is 2.66) to the total number of deviations from the median, $522 + 2.66 = 524.66$, which gives the total number of deviations about the true median.
11. Divide this sum by the total number of cases $\left(\frac{524.66}{289} = 1.816\right)$ to get the mean (average) deviation about the true median *EXPRESSED IN UNITS OF CLASS-INTERVALS.*
12. Multiply this mean deviation by the number of units in a class-interval to get *the mean deviation expressed in units of the original measures;* $1.816 \times 5 = 9.08$.

4. The standard deviation

Sigma as a measure of variability. It has been pointed out that there is a unit measure of variability called the standard deviation, sigma, (σ) that is coming into common use in educational measurement. It was stated there that if (on a symmetrical frequency distribution represented by the "probability curve"), a distance equal to the *standard deviation* is laid off on each side of the mean, and ordinates are erected from the base line to the curve, that between the curve, the ordinates, and the base line will be included 68.26 per cent of the measures represented by the total area. The derivation of the relation between sigma and the curve, the fact that it is a function of the curve, and hence may be used as a unit in laying off distances on the base line of the curve, the method of computing the percentage of cases under the curve, and between ordinates erected at stated distances of multiples or fractions of sigma from the mean: — these, and other points will all be cleared up in Chapter VII. It is our business here, however, to familiarize ourselves thoroughly with sigma as a measure of variability of a frequency distribution.

It was stated in the foregoing discussion that the computation of the mean deviation involves the arbitrary procedure of disregarding the signs of the deviations. The *standard deviation*, introduced by Karl Pearson in 1896, avoids this step by involving the *squaring* of all deviations from the mean. It differs from the mean deviation only in that feature. We may, therefore, define the standard deviation of a distribution, sigma (σ) as, *the square root of the arithmetic mean of the squares of the deviations from the average of the distribution.*

For the simple series: —

$$\sigma = \sqrt{\frac{\Sigma d^2}{N}}$$

where σ = the standard deviation, d = the deviation of any measure from the arithmetic mean, and N = the number of measures in the distribution.

For the frequency distribution: —

$$\sigma = \sqrt{\frac{\Sigma f d^2}{N}}$$

f representing the frequency of occurrence of the measures in any class-interval.

For approximate work with educational data it is practicable to take the deviations from either the arithmetic mean or the median. The natural average to use, however, is the arithmetic mean, for it is about this point in the distribution that the sum of the squares of the deviations is a minimum. (The mathematical theory underlying the derivation of refined statistical measures makes use of the principle that the sum of the squares of deviations should be a minimum.)

Computation of the standard deviation: (*a*) **data ungrouped.** The steps in the computation of the standard deviation, when the measures are ungrouped, are simply stated as given below. Each step is illustrated by data from Table 24, columns 4 and 5.

1. Compute the arithmetic mean of the series, 79.97.
2. Compute the deviation of each measure from the mean; 16.03; 15.03, etc.
3. Square each deviation; (use tables for squaring).
4. Find the total of such deviations, 256.9; 225.9, etc., = 2871.8.
5. Find the arithmetic mean of the deviations; = 110.45, this is σ^2.
6. Find the square root of the mean; 10.51.

(*b*) **frequency distribution.**

This is the standard deviation of the distribution, σ.

Just as with the computation of the arithmetic mean, the labor of computing the standard deviation by this long method is considerable. It may be very materially cut down by recourse to the short method, explained in connection with the arithmetic mean and the mean deviation. That method makes use of the principle that instead of computing the deviation in each case from the true mean or median, thus giving three- or four-place numbers in the later multiplication, as in Table 24, an assumed mean is taken (*anywhere in the distribution*), and the deviations are computed from this point. Furthermore, in the case of the frequency distribution, the assumed mean is taken at the mid-point of a class-interval, and the deviations are all laid off in units of class-intervals, thus reducing the arithmetic work to a minimum. To get the method before us clearly, we next list the steps in the computation that are necessary when the data are arranged in the frequency distribution. The data of Table 30 may be taken to illustrate each step in the procedure.

Steps in the procedure of computing the standard deviation by the short method. We may now summarize the steps in computing the standard deviation by the short method, as follows: —

1. Tabulate the frequency distribution.
2. Estimate the interval which contains the mean; select it near the point of heaviest concentration of the measures; *e.g.*, 41.0–44.99.
3. Tabulate the deviation in unit intervals, of the mid-value of each class-interval from that of the estimated mean. 1, 2, 3, etc.; −1, −2, −3, etc.
4. Multiply each frequency by its respective deviation; $f \times d$, 10, 9, 16, 14, etc.
5. Find the algebraic sum of such *fd's*; $\Sigma fd = 327 - 336 = -9$.
6. Divide Σfd by the number of cases. (This quotient is the correction "*c*" which (multiplied by the number of units in a class-interval) added (algebraically) to the estimated mean gives the true mean. This true mean does not necessarily have to be computed to get the standard deviation.) $c = -.03$.
7. Multiply each fd by d, its corresponding deviation, giving the column headed fd^2; 100, 81, 128, etc.

TABLE 30. DISTRIBUTION OF ABILITIES IN VISUAL IMAGERY OF 303 COLLEGE STUDENTS TO ILLUSTRATE THE COMPUTATION OF THE STANDARD DEVIATION BY THE SHORT METHOD

Class-interval	*Frequency* f	*Deviation* d	fd	fd^2
90.0–94.99	1	10	10	100
85.0–89.99	1	9	9	81
80.0–84.99	2	8	16	128
75.0–79.99	2	7	14	98
70.0–74.99	3	6	18	108
65.0–69.99	5	5	25	125
60.0–64.99	7	4	28	112
55.0–59.99	26	3	78	234
50.0–54.99	41	2	82	164
45.0–49.99	47	1	47	47
40.0–44.99	50	0	327	
35.0–39.99	32	−1	−32	32
30.0–34.99	31	−2	−62	124
25.0–29.99	18	−3	−54	162
20.0–24.99	16	−4	−64	256
15.0–19.99	11	−5	−55	275
10.0–14.99	3	−6	−18	108
5.0– 9.99	5	−7	−35	245
0.0– 4.99	2	−8	−16	128
	$N = 303$		−336 327	303)2527(8.34 $= S^2$
			303)−9(−.03	

$c = -.03 \quad 8.34 - .001 = 8.339 = \sigma^2$

$c^2 = \ \ .001 \quad \sigma = 2.88$ intervals

$\sigma = 2.88 \times 5 = 14.4$ actual units

8. Find the sum of the fd^2's; $\Sigma fd^2 = 2527$.
9. Divide the sum of the fd^2's by the number of cases to get S^2, $S^2 = 8.34$. It must be remembered that S^2 is the square of the "standard deviation" (technically called the *root-mean-square* deviation) *from any assumed mean*. The mean of the deviations about this assumed mean is obviously in error by an amount equal to the arithmetic mean of the difference of the positive and negative deviations. (If taken about

the true mean the difference should be 0.) In the same way, the arithmetic mean of the squares of the deviations is in error by an amount equal to the square of this difference or c^2.

10. Square the correction, giving c^2, .001.
11. Subtract[1] c^2 from S^2, giving σ^2. This standard deviation is expressed in units of class-intervals; $8.34 - .001 = 8.339$.
12. Find the square root of σ^2 to give σ, — *still in units of class-intervals;* 2.88.
13. Turn σ (as expressed in units of class-intervals) into a σ expressed in unit measures, by multiplying σ by the number of units in a class-interval. $2.88 \times 5 = 14.4$.

As a rough check on the numerical work, it is well for the student to remember that for fairly long symmetrical or moderately-skewed distributions a distance of $6 \times \sigma$ includes 99 per cent of the measures. Reference to this will often prevent gross errors. There is a specific use of this "inspectional" method in the determination of the value of the Pearson coefficient of correlation. This coefficient may be roughly estimated by inspection of the contingency table by noting the spread of the distribution. The extent of this spread may be estimated numerically by the use of the empirical rule above. This will facilitate the approximate determination of the correlation coefficient.

---

[1] The square of the correction is *always subtracted from* S^2 (the "standard deviation" about the assumed mean). The proof of this algebraically is adapted from Yule, G. U., *Introduction to the Theory of Statistics*, page 134, as follows: —

Let x be any variable.
Let M be the true mean value of x.
Let A be any assumed value of the mean value of x.
Let $Z = x - A$. This is the deviation of the mean value of x from A.

Now define any general root-mean-square deviation, S (*e.g.* σ), from the origin A, by

$$S^2 = \frac{\Sigma Z^2}{N}$$

σ, is then the root-mean-square deviation from the true mean. Now to find the relation between σ and the root-mean-square deviation from *any other* origin, —

$$\text{Let } M - A = c \text{ so that } Z = x + c$$
$$\text{Thus } Z^2 = x^2 + 2xc + c^2 \text{ or}$$
$$\Sigma Z^2 = \Sigma x^2 + 2c\Sigma x + Nc^2$$

Now Σx, the sum of the deviations, is equal to 0, and from above $\Sigma Z^2 = NS^2$.

Therefore

$$NS^2 = N\sigma^2 + Nc^2$$
$$S^2 = \sigma^2 + c^2 \text{ or } \sigma^2 = S^2 - c^2.$$

Advantageous properties of the standard deviation. Using the list of desirable properties of the various means as given above as a criterion for establishing the value of σ, we may say that it ranks high as a measure of variability because:

1. It is numerically defined.
2. It is based on all the measures.
3. It is easily calculated.
4. It is susceptible of algebraic treatment (*e.g.*, it can be shown that the square of the σ of a distribution is equal to the arithmetic mean of the squares of the σ's of component parts of distribution.)
5. It can be shown by the theory of errors and sampling, that it is the measure of variability least affected by fluctuations of sampling.
6. Its computation aids the determination of the Pearson coefficient of correlation.
7. It is convenient because of the necessity of obtaining a measure which will vary with the variability of distribution, and squaring deviations is the simplest method of eliminating signs.
8. It bears a convenient relationship to the normal or probability curve, in that it is the distance from the mean to the point of inflection of the curve, *i.e.*, to the point of change of curvature. This will be made clear in the discussion of the graphic representation of measures.

The general rule may be laid down that the arithmetic mean should ordinarily be used as a type, or average, and the standard deviation (deviations all measured from the arithmetic mean) should be used as a measure of variability in all fairly long and symmetrical distributions met in educational research.

II. MEASURES OF RELATIVE VARIABILITY

Types of such measures. In the foregoing pages we have pointed out the principal methods of representing the *absolute* degree of variability of a given frequency distribution. Measures of variability are principally of value, however, in comparing one distribution with another. It is clear that standard deviations, mean deviations, or probable errors, *in order to be comparable, must be measured about averages of approximately the same absolute value.*

We must recognize two distinctly different kinds of variability in our measurements: (1) that in which two distributions are compared that have the same unit of measurement, — *e.g.*, salaries of teachers in two cities, achievement of two classes of pupils in a given standard test, or percentile distribution of municipal expenditures to various city departments; (2) that in which two distributions are to be compared in which the units of measurement are entirely different, — *e.g.*, the achievement of a class of pupils in handwriting as measured by the Thorndike Scale (units of 1 ranging from 4 to 18) with that of another class as measured by the Ayres Scale (units of 10, ranging from 20 to 90); or the salaries of teachers compared with their years of experience. Both types of variability need to be discussed here.

1. Unit of measurement the same. In order to secure comparable measures of variability it is not sufficient that the unit of measurement be the same. Examination of the data of Table 31 shows that it is not proper to compare the *absolute* variability of municipal expenditures for schools with that of expenditures, say, for the health department. If we used the *absolute* variability as shown by the respective mean deviations we would conclude that cities are nine times as variable in their expenditures for public school purposes as for public health purposes. This does not at all agree with the conclusion to be made from the logic of the situation, which is, that variability in expenditures for schools must be relatively small, and that for health relatively large.

Examination of these data shows us that the *gross absolute variability* is directly contributed to by the absolute value of the average from which we take our deviations, — in other words, by the magnitude of the units included in that portion of the scale covered by our distribution. Obviously, an absolute measure of variability must be many times larger when computed from an average of 32.30 than when computed from an average of 1.40.

2. Unit of measurement different. Similarly, it seems clear that we cannot compare the absolute variability of a group when measured with one unit, say salaries of teachers, in dollars, with that of the same or another group when measured in terms of an entirely different unit, say, their years of experience in teaching. In the former case, our range might extend from $250 to $1175, the average be $640, and the *M.D.* perhaps be $150. In the latter case, the respective measures might be, — range, 1 to 37, average 9, *M.D.* 5. Thus it is clear that we need a measure of RELATIVE *variability* to cover these two cases. Evidently we must conclude that only when two distributions give about the same average, and cover about the same portion of the scale, are their measures of absolute variability directly comparable.

The Pearson coefficient. To take account of the relative magnitude of the average and of the units on the scale the suggestion has been made by Pearson that we find the ratio of the measure of absolute variability (σ, $M.D.$, Q, or $P.E.$), to the average from which the deviations were taken (arithmetic mean or median). Expressed in algebraic form this is

$$V = \frac{100\,\sigma}{M}$$

called by Pearson the *coefficient of variation.* A measure of this type is evidently independent of the magnitude of the

TABLE 31. AVERAGE PERCENTILE PAYMENTS FOR GENERAL AND MUNICIPAL SERVICE — FISCAL YEARS 1902 AND 1903 — CITIES BETWEEN 25,000 AND 50,000 POPULATION *

Municipal activities	(1) *Median Md*	(2) *Average deviation A.D*	(3) *Thorndike's coefficient of variability*	(4)† *Pearson's coefficient of variability*
General Administration..	8.08	1.54	.54	.19
Police Department.......	8.16	1.74	.609	.21
Fire Department.........	9.98	2.58	.817	.26
Health Department......	1.40	.747	.633	.53
Charities and Corrections	3.02	2.98	1.71	.99
Public Highways.........	8.19	2.52	.908	.31
Street Lighting..........	6.43	1.84	.725	.29
Public Sanitation........	3.67.	1.78	.927	.49
Schools..................	32.30	6.67	1.175	.21
Libraries	1.14	.56	.524	.49
Public Recreation........	.61	.642	.814	1.05
Interest on Debt	12.50	5.75	1.62	.46

* Adapted from Elliott, *Some Fiscal Aspects of Education*, p. 83.
† Column 4 added by the writer for comparative purposes.

units on the scales of the two distributions. In using it, one is merely finding the per cent (if he multiplies $\sigma \times 100$) that the absolute variability is of the average from which the deviations are computed. It is clear that the same type of measure could be obtained by dividing the quartile deviation or the mean deviation by the median. To do this in the case of the data in Table 31 gives the coefficients in column 4. According to these, the item of expenditures for "schools" is among the least variable, ranking with those for police, fire, general administration, etc.; and public recreation, charities and corrections, and health are among the most variable. These statistical conclusions clearly check those inferred from our logical analysis of the situation, and aid us by enabling us to speak in fairly definite terms. Expressed in another way, cities agree much more closely in their expenditures for the old established departments of schools, fire, police, general administration, highways, than they do for the newcomers in the field of municipal administration, — public recreation, health, charities, etc.

Thorndike, however, proposes another empirical measure of relative variability, choosing to divide the measure of deviation *by the square root of the average;* thus: —

$$V = \frac{100\ M.D.}{\sqrt{Median}}$$

The results of using this measure instead of the direct percentage of deviation and average, are, according to Thorndike, "more in accord with both theory and facts." The data of column 3, Table 31, were originally computed by Elliott with the use of Thorndike's coefficient, his interpretations and conclusions being determined by the relative size of the coefficients. He says, for example: "From these coefficients it is justifiable to say that the expense for libraries and that for general administration seems to be least subject to the influence of those conditions likely to produce variability, while the expense for charities and corrections, interest on the debt, and schools, possess, in the order named, the largest degree of variability." It is plain that these conclusions are quite the reverse of those deduced above from the logic of the situation, and which are also obtained from the use of the Pearson coefficient. Furthermore the taking of a coefficient containing a root or power of the mean used as base makes the coefficient very unstable when applied to problems in which that measure varies widely in magnitude. Contrast, for example, the effect of having a base (median) closely approximating 1 (such as health expenditure above) in which the square root of the base varies but little from the size of the base itself, with the case of schools, in which the base is 32, the square root of which becomes 5.657. To get the full effect of the liberties that one takes with ratios of this type let us illustrate by a simple problem.

Suppose, in a given distribution, a median to have been computed of 10 feet, with a corresponding mean deviation of 3 feet. The Pearson coefficient of variation = $\frac{3}{10}$, the Thorndike,. $\frac{3}{\sqrt{10}}$. Now express the same measures in inches, getting, median 120 inches, mean deviation 36; Pearson coefficient, $\frac{36}{120}$ or $\frac{3}{10}$ as before ; Thorndike coefficient, $\frac{36}{\sqrt{120}}$. The manipulation necessary to get the latter result makes $\frac{3}{\sqrt{10}} = \frac{36}{\sqrt{120}}$, or a coefficient of variability of .949, on the same measures becomes, by merely refining the unit of the scale, 3.28. The writer's experience leads to an acceptance of Pearson's coefficient as a helpful device in roughly comparing the spread of two distributions.

III. MEASURES OF SKEWNESS OR LACK OF SYMMETRY IN DISTRIBUTIONS

In the previous discussion of the treatment of frequency distributions, constant emphasis has been laid on the symmetry of the distribution in question. It was said repeatedly that certain measures could be used (*e.g.*, the probable error), if the measures were distributed approximately symmetrically about the average of the group. Statisticians have thus faced the need of devising a single coefficient to express the degree to which the distribution is "*skewed*," or the degree to which it *lacks symmetry*. It is clear that this coefficient must be independent of the magnitude of the scale units, and we wish to represent it as a single number. Examination of Diagrams 20 and 21 will show the reader that a measure could be built up by expressing the relation between the mean, the median, and the mode in some fashion. In the perfectly symmetrical distribution, Diagram 28, they all coincide. With partially skewed distributions the mean, mode, and median stand in a somewhat constant relation to each other, such that the median lies at a point approximately one third of the distance from the mean toward the mode. Reference to the discussion of relative variability in the previous section will remind the reader that this relation between mean, mode, or median should be measured in terms of a unit of deviation. To satisfy these various criteria, Pearson has suggested the use of the following measure of skewness: —

$$Skewness = \frac{Mean - Mode}{\sigma}$$

Since the true mode is very difficult to determine, we might use the approximate formula for it by recalling the relation between mean, median, and mode, getting: —

$$Skewness = \frac{3\ (mean - median)}{\sigma}$$

Yule has suggested that an approximate measure of the same type might be built up by finding the difference between the two middle quartiles, *provided we measure this difference by its ratio to some standard measure of variability*, measured in the same units, — for example, the quartile deviation Q. In those cases where Q is used as a measure of variability,

$$Skewness = \frac{(Q_3 - Md) - (Md - Q_1)}{Q} = \frac{Q_1 + Q_3 - 2Md}{Q}$$

ILLUSTRATIVE PROBLEMS *

1. Find the quartile deviation, the mean deviation, and the standard deviation for each of the frequency distributions reported in the "illustrative problems" of Chapter V.

2. Find the coefficient of variation for each of these problems by the Pearson formula and by the Thorndike formula.

GIVEN FOR FOUR DISTRIBUTIONS

	Distribution A *Number of words read per second*	*Distribution B* *Percentile marks given pupils in drawing*	*Distribution C* *Number of arithmetic problems solved per minute*	*Distribution D* *Marks given pupils in mathematics*
Arithmetic Mean ...=	3.9	77.8	12.4	76.9
σ.........=	1.4	19.3	2.9	11.3

Questions: 1. In which of these distributions is the variability greatest?......2. Which may be compared directly by means of their measures of absolute variability?......3. Why?

---

* Quoted from Rugg, H. O., *Illustrative Problems in Educational Statistics*, published by the author to accompany this text. (University of Chicago, 1917.)

CHAPTER VII

THE FREQUENCY CURVE

THIRD METHOD OF DESCRIBING A FREQUENCY DISTRIBUTION

Summary of preceding work. We have been continually trying to find the best methods of describing a frequency distribution. We have tried the use of the "range," or the distance on the scale between the lowest and highest values. It was noted that this number depends solely on two values of measures which are subject to great fluctuation, namely, the largest measure and the smallest measure. We have tried to typify distributions by various "averages," but it was shown again that either the arithmetic mean, median, or mode can but *partially describe* the distribution. In other words, two distributions *may vary widely* in the way in which the measures are concentrated or scattered along the scale, at the same time that they present exactly the same "average." So we have turned to the method of variability, and have discussed the use of measures to represent the amount of this "scattering" or "bunching" of measures. It was shown that a fairly adequate numerical representation of the two distributions in question could be obtained by giving both the average and the variability (*e.g.*, by the arithmetic mean and the standard deviation, or the median and the mean deviation, or the median and the quartile deviation, etc.). These could be supplemented, in cases where the units of the scale of the two distributions were different, by a coefficient of relative variability.

Our sole aim in treating educational data by any of these devices is to organize a complex mass of material in such a way as to *facilitate clear educational interpretations*. It seems quite clear that the mind finds it difficult to deal with whole frequency distributions, or with the original ungrouped measures themselves. The "average" and measure of "variability" help to condense the material and aid in interpretation. It was pointed out in Chapter IV that thorough use can be made of such measures only by the most experienced manipulator of statistical methods; that the student needs other methods of representing facts. It was shown that probably the greatest aid to sound interpretation of statistical data can come from the graphic presentation of the facts in question.

Smoothing frequency polygons to approximate ideal "distributions." It is suggested that at this point the student review the discussion of methods of plotting educational data in the form of the frequency polygon and the column diagram (Chapter IV). It was emphasized there that, although we actually deal with but a small proportion of the total population of measures similar to those in question, our desire for educational interpretations of the data leads us to speak in terms of *the frequency curve* which is believed to typify the law underlying our distribution. To be concrete: —

Diagram 30 reveals clearly that it is drawn to represent a *limited* number of measures. If we had had an infinite number of measures, and the size of the class-intervals had been "very small," the polygon of Diagram 30 would have become a "smooth" curve, perhaps somewhat like those in Diagrams 30 and 31. The matter can be more clearly explained from Diagram 31.

Assume that we can refine our measuring scale so as to get class-intervals of, say, tenths or hundredths of a unit, instead of 5 units. Furthermore, assume that we increase the number of measures from 303 to some relatively large

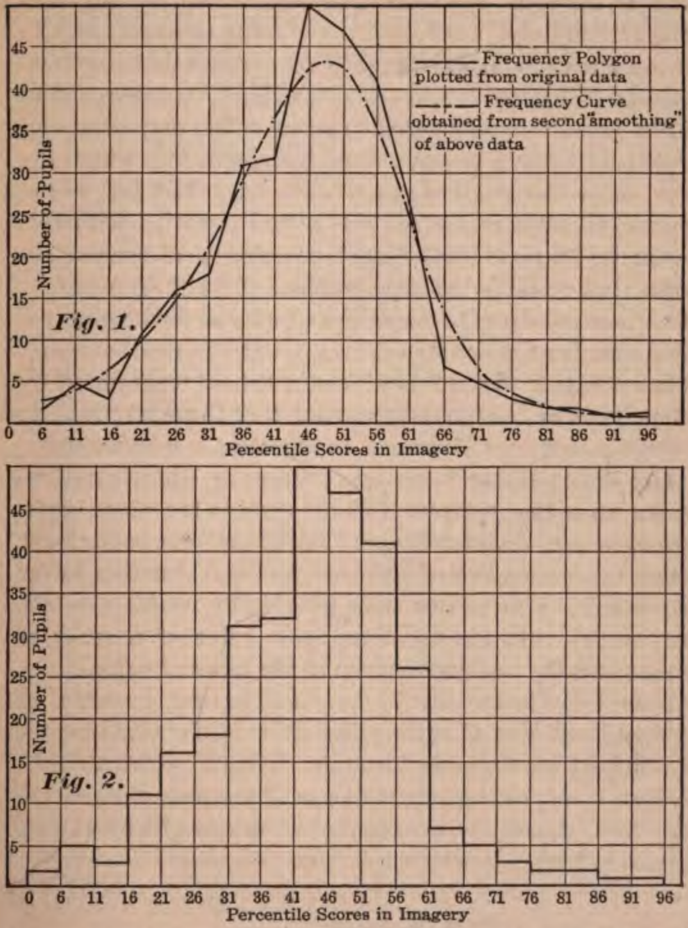


Diagram 30. Frequency Polygon (Fig. 1) and Column Diagram (Fig. 2) to represent Distribution of Abilities of 303 College Students in Visual Imagery

(Data in Table 30.)

number, say 3000, or 30,000. The base of each rectangle becomes "infinitely" small, and the number of cases tends to be more *continuously scattered.* Thus we find that our "rectangular histogram" approaches "as a limit," some smoothed curve, perhaps having specific mathematical properties and capable of leading to generalized interpretations which the very particularized histogram does not. We say, — the column diagram represents the actual situation with this particular "sample of 303 cases"; the smoothed curve represents what would be the most probable value of the measures at various points on the scale if we took the entire group of measures (from which we actually have but a small sample). It is very clear that a "law" could not be represented by the polygon or column diagram, but that the most probable definite curve must be sought to represent it adequately.

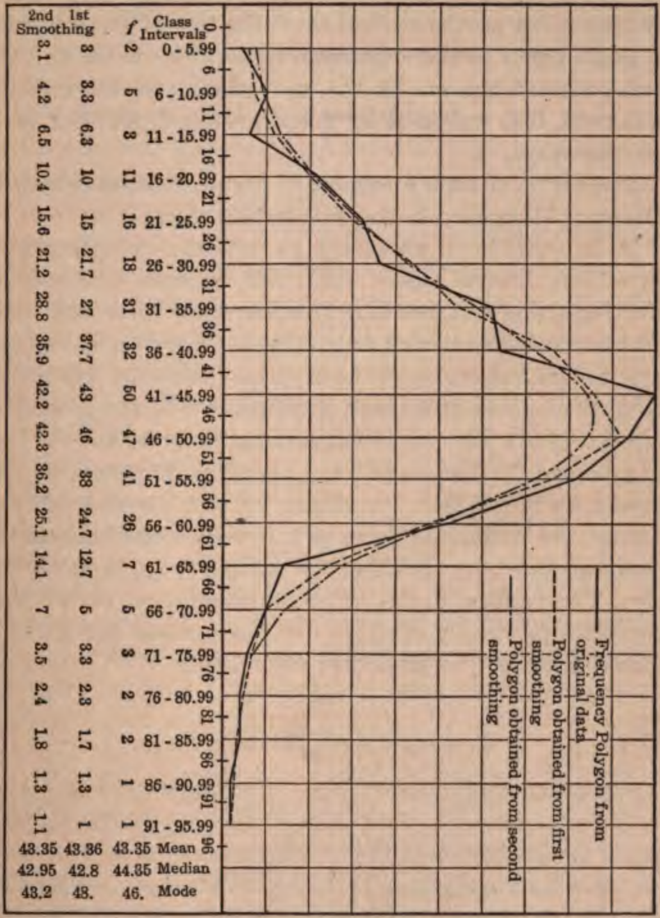
The "smoothing" process. Since in educational research we cannot work with *all* the cases in the entire population, we may be interested in "*smoothing*" our polygons or column diagrams to approximate the ideal situation as far as possible. This can be done roughly by working on the assumption that the most probable value of a series of measures is the arithmetic mean of the series of values.

This hypothesis can be applied to our problem of "smoothing" by taking the arithmetic mean of small groups of adjacent measures on the scale. Thus if we let A, B, C, D, E, F, etc., be the actual values of the midpoints of the intervals, we may average the number of cases found at each three adjacent points by the formula: —

$$\text{Smoothed value of } A = \frac{2A + B}{3}$$

$$\text{Smoothed value of } B = \frac{A + B + C}{3}$$

$$\text{Smoothed value of } C = \frac{B + C + D}{3}$$



2nd Smoothing	1st Smoothing	f	Class Intervals
3.1	3	2	0 - 5.99
4.2	3.3	5	6 - 10.99
6.5	6.3	3	11 - 15.99
10.4	10	11	16 - 20.99
15.6	15	16	21 - 25.99
21.2	21.7	18	26 - 30.99
28.8	27	31	31 - 35.99
35.9	37.7	32	36 - 40.99
42.2	43	50	41 - 45.99
42.3	46	47	46 - 50.99
36.2	38	41	51 - 55.99
25.1	24.7	26	56 - 60.99
14.1	12.7	7	61 - 65.99
7	5	5	66 - 70.99
3.5	3.3	3	71 - 75.99
2.4	2.3	2	76 - 80.99
1.8	1.7	2	81 - 85.99
1.3	1.3	1	86 - 90.99
1.1	1	1	91 - 95.99
43.35	43.36	43.35	Mean
42.95	42.8	44.35	Median
43.2	43.	46.	Mode

DIAGRAM 31. SAME DATA AS IN DIAGRAM 30. COMPARISON OF "ACTUAL" FREQUENCY POLYGON WITH RESULT OF FIRST AND SECOND "SMOOTHINGS"

etc., throughout the series. It is seen that the "true" value of each point on the scale (except the two extreme values) is taken equal to the arithmetic mean of its value and the two adjacent values. In the case of each of the extreme measures, it is weighted by 2 and averaged with the adjacent measure.

The result of such a scheme of approximation is seen in Diagram 31, applied to the distribution of Table 30.

It is sometimes necessary to repeat the process of smoothing several times. This will be true especially in those distributions revealing sharp irregularities. It is clear that in most educational distributions these irregularities or "peaks" in the curve will be explained either by scarcity of number of cases, or by lack of refinement in the process of measurement. The numerical and graphic results of the first and second "smoothings" are shown in Diagram 31. It should be noted that smoothing by this method will not change the arithmetic mean of the whole distribution. On the other hand, it may affect the median or mode considerably. The results of the different smoothings reveal that beyond a particular repetition of the process but little is gained in the way of smoothed refinement.

I. IDEAL FREQUENCY CURVES

School-marking distributions. Fundamentally necessary to the advancement of all phases of school practice is adequate knowledge about the intellectual and physical capacities of school children. The design of a course of study, planning of teaching methods, adapting of all such phases of school machinery as grading and classification of children, their promotion from grade to grade, marking systems, — all these questions rest back upon the possibility of being able to picture completely the distribution of abilities in our school population. For example, the design of a marking system, or of standardized tests for the measurement of ability in any subject of study, must rest upon clear-cut hypotheses as to the distribution of ability in the school population in question.

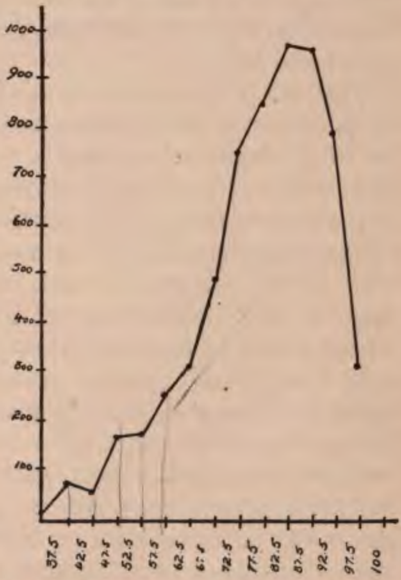


DIAGRAM 32. DISTRIBUTION OF 5714 MARKS GIVEN IN PLANE GEOMETRY IN FIFTEEN HIGH SCHOOLS

Compare this Diagram with Diagram 24.

Let us take a concrete example, using data in the situation represented by Diagram 32; this gives the actual distribution of 5714 pupil marks in 15 high schools in plane geometry. The curve shows that over 30 per cent of the pupils were classed as being 90 per cent in ability or above, *i.e.*, in the top fifth of 5 groups of ability. We are at once skeptical of the accuracy with which the teachers have judged the abilities of these pupils, all the more so when we note that the curves are concentrated at 75 and 90 and when we find that these points on the scale represent the passing and exemption marks respectively.

On comparing our data with those in Diagram 24 we are convinced that the marking machinery does not represent accurately the abilities of pupils. Here, we note that, as the result of careful testing of intelligence, arithmetical ability, stature, and other anthropometrical measurements, the top fifth of our pupils is surely not more than 6 to 10 per cent of the total group. Certainly there is no reason to believe that even our high-school population is so badly "skewed" in ability that nearly one third falls closely together in the top fifth of the scale.

Now, the administration in this particular system has recognized recently that its marking is not fitted to the capacities of pupils, and has faced the very real question: "With what relative frequency should pupil-ability be distributed in the various fifths of the marking scale? What per cent of the total group actually merit A, B, C, D, E?" To answer this question fully, this superintendent needs detailed objective evidence on the distribution of similar high-school pupils in large numbers. If he could secure it he would be perfectly justified in educating his teaching staff to the point where it would measure pupils' abilities roughly in accordance with this objectively-obtained distribution.

The distribution of human traits. Complete figures on the abilities of high-school pupils are lacking, but he has available many measurements on human intelligence, various mental traits, and a vast amount of evidence concerning the distribution of anthropometrical measurements on human beings. The student will be interested to note with what striking regularity they resemble a fairly symmetrical curve. In *all* such distributions, the measures are largely concentrated very near the middle of the scale. Furthermore, they shade off in both directions from the middle high point, — the mode, — somewhat symmetrically. The student will note, furthermore, that in the case of those traits which are more subject to refined measurement, — *e.g.*, heights of men, strength of grip, cephalic index, chest measurement and other physical measurements, and fairly refined psychological measurements, the curves the more closely approximate symmetry. In addition, we see that in those cases where very large numbers of measurements have been taken, as in Diagram 24, Fig. 4, heights of men, the curve strikingly approaches this symmetrical type.

A century ago, the regularity of this accordance of the distribution of human traits with definite symmetrical curves was noted by various observers. Quetelet, the Belgian scientist, made many such measurements, and early called attention to the recurring conformance of the shape of the curve of human measurements to the chance polygon got by plotting the coefficients of the separate terms in the binomial expansion. Especially close is the "fit" in the case of such physical measurements as stature and girth of chest.

Laws of nature show continuous distributions. With the agreement upon the shape of the distribution curve of human traits there came a recognition of the need for the definite establishment of ideal curves which could be used in the case of interpretation of fairly limited numbers of observations or measurements. Science demanded a means of generalization — a method of expressing "the law." More and more they commented on the fact that laws of nature, as generalizations based on human experience, were interpreted only in terms of continuous distributions. The distribution of human measurements was checked further by the distribution of "errors of observation" in refined measurement, — *e.g.*, astronomy, surveying, etc. The plotting of such refined measurements gave a distribution resembling, in a rather striking way, the shape of the curve of distribution of human traits, concentrated near a mode about the middle of the range, sloping off quite symmetrically in both directions, and showing relatively few cases at the extremes. If the errors be plotted with the error 0 at the middle and positive and negative errors plotted on either side of this point, this may be interpreted partially by saying: *first*, that very small errors are most common (the error "zero" is really most common); *second*, that positive and negative errors are about equally frequent; and *third, that very large errors do not* occur. This may be illustrated by a brief quotation from Merriman's *Method of Least Squares* (p. 13): —

> For instance, in the *Report of the Chief of Ordnance for 1878*, Appendix S′, Plate VI, is a record of one thousand shots fired deliberately (that is, with precision) from a battery-gun, at a target two hundred yards distant. The target was fifty-two feet long by eleven feet high, and the point of aim was its central horizontal line. All of the shots struck the target; there being few, however, near the upper and lower edges, and nearly the same number above the central horizontal line as below it. On the record, horizontal lines are drawn, dividing the target into eleven equal divisions; and a count of the number of shots in each of these divisions gives the following results: —

In top division...............	1 shot
In second division............	4 shots
In third division.............	10 shots
In fourth division............	89 shots
In fifth division..............	190 shots
In middle division............	212 shots
In seventh division...........	204 shots
In eighth division............	193 shots
In ninth division.............	79 shots
In tenth division.............	16 shots
In bottom division...........	2 shots
Total.................	1000 shots

> It will be observed that there is a slight preponderance of shots below the center, and there is reason to believe that this is due to a constant error of gravitation not entirely eliminated in the sighting of the gun.
>
> The distribution of the errors or residuals in the case of direct observations is similar to that of the deviations just discussed. For instance, in the *United States Coast Survey Report for 1854* (p. 91) are given a hundred measurements of angles of the primary triangulation in Massachusetts. The residual errors (art. 8) found by subtracting each measurement from the most probable values are distributed as follows: —

Between + 6.0 and + 5.0	1 error
Between + 5.0 and + 4.0	2 errors
Between + 4.0 and + 3.0	2 errors
Between + 3.0 and + 2.0	3 errors
Between + 2.0 and + 1.0	13 errors
Between + 1.0 and 0.0	26 errors
Between 0.0 and − 1.0	26 errors
Between − 1.0 and − 2.0	17 errors
Between − 2.0 and − 3.0	8 errors
Between − 3.0 and − 4.0	2 errors
Total	100 errors

Here also it is recognized that small errors are more frequent than large ones, that positive and negative errors are nearly equal in number, and that very large errors do not occur. In this case the largest residual error was 5.2; but, with a less precise method of observation, the limits of error would evidently be wider.

The axioms derived from experience are, hence, the following: —

1. Small errors are more frequent than large ones.
2. Positive and negative errors are equally frequent.
3. Very large errors do not occur.

II. The Normal Probability Curve

Resemblance of actual distributions to "chance" distributions. Enough has been said to point out the very practical need in all the sciences for a distribution curve, from which generalizations could be made. It was early recognized by these workers that their distributions resembled in a striking way the shape of the frequency polygons obtained by plotting the frequency of various "chances." Since the manipulation of the mathematical properties of the distribution of "chance" leads to the ideal curve which we are seeking, we shall next turn to a very elementary discussion of "chance" and the "probability" curve. Before doing so, let us state clearly the ultimate goal of the student of educational problems, in seeking an ideal curve against which he can check his actual distribution and from which he can generalize his experience. Expressed briefly, it is this: —

1. Knowing that human traits distribute closely enough for practical purposes in accordance with a particular ideal distribution, we wish to be able to locate easily the proportion of our total group (assuming it to be reasonably large) that should fall between any two points on the scale of our measurements. Concretely, our superintendent named above, wishes to know about how large a group of his pupils should get A's, B's, C's, D's, and E's. He also wants the process of this determination reduced to a minimum of arithmetic labor. In other words, our theory should lead to the preparation of tables by which the student can compare, easily and yet accurately, actual with ideal distributions.

2. Another important goal of the student of education in dealing with "probability" is found in connection with his very real need for being able to establish the reliability of his data. He is measuring a relatively small "sample" of the total group, and has computed averages, measures of variability, and perhaps of relationship. What dependence can be placed on the representativeness of the small sample? If he took other succeeding samples, would his measures of type and variability be practically what he has already found? Or can he feel assured that they would fluctuate much, and hence that from his data he can make no sound interpretation? It should be stressed here that adequate educational interpretations of the results of research must rest upon careful determination of the reliability of measures which have been computed. These two important needs of the student of education reveal the need for carefully acquainting ourselves with the way in which the "probability" distribution is found.

We have pointed out that human traits are "combinations" and include many "arrangements" of a vast number of separate causes which may be assumed to be independent of each other. In deriving a theoretical curve of distribution for a set of many independent causes, we must recall the mathematical result obtained by combining and arranging such groups of causes. It is of interest to note that the results of such combinations accord so closely with certain mathematical schemes; namely, those of permutations and combinations, which, working under the laws of probability, may be studied and whose conclusions may be applied to the interpretation of our data.

We shall next show the resemblance between the results of combining various arrangements of large numbers of independent causes and the straight mathematical theory of permutations and combinations. This leads to a statement of principles of GROUPING called combinations; and arrangements of same group or combination called permutations.

Use of permutations and combinations. From our elementary algebra we will recall that, with a given number of things we can make only a definite number of groupings or "combinations," each combination of things being different from any other. For example, let *a*, *b*, *c*, *d*, represent four things. We may make four, and only four different combinations of these four things when we take them three at a time, namely: —

abc, abd, acd, bcd.

If we take but two at a time we can make six, and only six different combinations, thus: —

ab, ac, ad, bc, bd, cd.

If we take four at a time, but one combination is possible, *abcd*. Now, with each of these combinations we may make two or more *arrangements* or *permutations*. The permutation is *determined by the order* in which the things stand. For example, with any such combination as *abc*, we may make 6 permutations: —

abc, acb, bac, bca, cab, cba.

Each thing here is combined with each remaining pair of things.

It is seen that the number of arrangements of n things (4) taken r (2) at a time is $n(n-1)$; *i.e.*, (4.3, or 12):

ab	*bc*	*cd*	*da*
ac	*bd*	*ca*	*db*
ad	*ba*	*cb*	*dc*

Take three at a time: —

abc	*bcd*	*cda*	*dab*
acb	*bdc*	*cad*	*dba*
abd	*bda*	*cdb*	*dbc*
adb	*bad*	*cbd*	*dcb*
acd	*bca*	*cba*	*dca*
adc	*bac*	*cab*	*dac*

$n(n-1)(n-2)$, or (24), and so on. The number of permutation of n things, taken r at a time is, therefore, —

$$n^{P}r = n(n-1)(n-2)\,\ldots\ldots\ldots\,(n-r+1).$$

Thus, since with any given combination of, say, r things, we can combine every thing with every remaining group of things, we can make factorial r permutations of things from the combinations. (Factorial r is written $r!$ or r and means $1, 2, 3, 4 \ldots\ldots r$.)

Therefore, as we take one combination after another of r things, with each combination we can make $r!$ permutations. Hence, the total number of permutations of n things taken r at a time is equal to the number of *combinations* of n things taken r at a time, multiplied by $r!$, or

$$n^{C}r^{r!} = n^{P}r\,; \text{ or, } n^{C}r = \frac{n^{P}r}{r!}$$

But, $$n^{P}r = n(n-1)\,(n-2)\ldots\ldots\ldots(n-r+1)$$
$$\therefore n^{C}r = \frac{n(n-1)\,(n-2)\ldots\ldots\ldots(n-r+1)}{r!}$$

We have said that "law" is but man's generalization from his experience. We are interested in seeing now in what way he can check his experience against regularity of mathematical order. The above formula for the number of combinations of n things taken r at a time now enables us to foretell, in the case of a given number, n, of independent events working under ideal conditions, what is the probability of a stated number of them, r, happening or failing to happen. To illustrate the operation of the principle let us take the case of coin tossing, assuming a coin to be a homogeneous disc and equally likely to fall heads or tails. Suppose we throw out four coins at random on a table. According to the law of combinations and permutations what should be the number of heads or tails turning up when r takes values of 0, 1, 2, 3, 4? There is now a total of 16 possible arrangements of heads and tails. Taking four at a time, say all heads or all tails, we can make but *one* possible combination, n_1, n_2, n_3, n_4; taking three at a time, say three heads and one tail, or vice versa, we can make *four* combinations: *e.g.*, —

$$n^{C}r = \frac{n\,(n-1)\ldots.(n-r+1)}{r!}\cdot \text{ or } \frac{4\cdot 3\cdot 2}{1\cdot 2\cdot 3} = 4.$$

Taking two at a time, two heads and two tails, we can make, $\frac{4\cdot 3}{2\cdot 1} = 6$ combinations. Since each time we throw out 4 coins, it is possible to make these combinations of heads and tails, we can infer that, should we continue to throw, we ought in the long run to stand a chance of getting various combinations of heads and tails in about the ratio of 1, 4, 6, 4, and 1. Now if we plot a polygon, making the heights of the ordinates equal, to scale, to these figures, we note that we have a symmetrical polygon, with, it is true, but five ordinates. Let us take a larger number of cases, or coins, say 7. Now "the chance of getting all heads," *i.e.*, the number of combinations that it is possible to make of 7 things, taking 7 at a time ($n = 7$, $r = 7$) is 1; the chance of getting 6 heads and one tail at a time is 7, the number 5 at a time, 21, the number 4 at a time, 35, etc. Thus we find that in the long run, our "chances" ought to be about 1, 7, 21, 35, 35, 21, 7, 1. Plotting these "chances" we find a polygon, with more ordinates, a flatter slope to its sides, but still symmetrical in shape.

Probability. Our discussion has now turned to the "chance" of this or that happening or not happening. It is possible then to extend our discussion in the form of general statements of probability, and thus establish an expression for the probability of any number of events happening or failing. To do that we must make clear what we mean by *probability* and establish certain fundamental principles. In defining probability we must recall that we are but trying to *idealize our actual experience* in order that we may establish what would be the most probable condition, in case our actual data could be made infinitely extensive.

To take a familiar actuarial illustration first: what is the probability that a particular child will not live to be 21 years of age? We are forced to turn at once to the actual experience of the human race under similar conditions. That is, we will find out what proportion of children actually have not lived to be 21, say 20 per cent or 1 out of every five. We idealize this experience by saying that since 1 in every 5 of a very large number of children fails to live to the age of 21, the *probability* that a child will fail to do so is 1/5. *Probability*, then, *is evidently to be defined as the ratio* between the occurrence of a particular event and the very large group of events of which it is a part. Or, expressed in another way, it means a number less than 1 — taken to represent the ratio of the number of ways in which an event may happen to the total number of possible ways, — *each of the ways being supposed equally likely to occur.*[1]

For example: if we toss a coin there are two possible ways in which it may come down, heads or tails. Hence the probability of its coming down heads is $\frac{1}{2}$, and of its coming down tails is $\frac{1}{2}$. The sum of the probabilities is of course unity, the mathematical symbol for *certainty*. For example, if the probability of hitting a target is 1/5000, the probability of not hitting it is 4999/5000.

Now, if an event may happen in *different independent ways*, the probability of its happening in either of these ways is the *SUM of the separate probabilities*. To illustrate: if we put into a bag 12 green, 18 red, and 19 black balls, and draw out a ball, the probability that it will be green is 12/49 (the total number of balls is 49, and there are 12 green ones); that it will be red, 18/49; and that it will be black 19/49. But the probability that it might be either black or green will be

$$\frac{19}{49}+\frac{12}{49}=\frac{31}{49}$$

and the probability that it might be either black, green, or red, is

$$\frac{19}{49}+\frac{12}{49}+\frac{18}{49}=1.$$

If we let P represent the probability of an event happening, and Q that of its not happening, then

$$P=1-Q \text{ or } P+Q=1.$$

Probability in educational research. Now, in our research we are dealing with "compound" events; *i.e.*, those produced by the concurrence of a very large number of causes,

---

[1] The writer has adapted to his uses here, Merriman's discussion of probability and the binomial expansion in *Methods of Least Squares*, pp. 6–10.

assumed to be independent of each other. For example, "arithmetical ability" in a particular individual, may be said to be a complex resultant of a very large number of causes, *e.g.*, those due to hereditary capacity, physical conditions of growth, and conditions of home and school training; *e.g.*, absence from or regularity in school, outside activities, etc. We cannot isolate the specific unit causes, so hopelessly are they tangled up, but we can measure the *effect of the combination* of this vast number of separate causes by the objective evidence; *i.e.*, we measure the resultant human trait called, for convenience, "arithmetical ability." Now it is a safe assumption that these many separate causes are independent of each other, — at least they are not related in any definite way. Human events thus are assumed to be "compounds," analogous in their determination to compound "chance" events of an ideal nature. That they show distributions of somewhat similar shape is very evident from our foregoing discussions. We need statements, therefore, for the generalization of such compound events.

What is the probability of the happening of a particular *compound event?* The answer must be, — *the product of the probabilities of the happening of the separate independent events.* For example: if one of two bags contains 8 black balls and 9 red balls, and the other contains 3 black balls and 11 red balls, the probability of drawing 2 black balls in 1 draw from each bag $= \frac{8}{17} \times \frac{3}{14}$. In the same way we may extend this to any number of events. $P_1P_2P_3P_4$ is the probability that all of four events will happen, and $(1 - P_1)\ (1 - P_2)\ (1 - P_3)\ (1 - P_4)$ is the probability that all will fail. Thus $P_1\ (1 - P_2)\ (1 - P_3)\ (1 - P_4) =$ the probability that 1 will happen and 3 will fail, etc.

Probability expression. We are now in a position to establish an expression for the probability of any number of events happening or failing. Assume "n" events, and assume that $P + Q = 1$. Then,

1. From the above, the probability that *all* events will happen is $P_1P_2P_3 \dots P_n = P^n$.

2. The probability that 1 assigned event will fail and $(n - 1)$ happens is $P^{n-1}Q$. Since this may happen in "n" ways, the probability that 1 will fail and $(n - 1)$ happen is $np^{n-1}q$.

3. The probability that 2 assigned events will fail and $(n - 2)$ happen is $P^{n-2}Q^2$; since this may be done in $\frac{n(n-1)}{1.2}$ ways (because the coefficient $= \frac{n(n-1)}{r!}$ the probability that 2 events out of the total will fail is $\frac{n(n-1)}{1.2} P^{n-2}Q^2$).

4. The probability that 3 assigned events will fail and $(n - 3)$ happen is $P^{n-3}Q^3$. Since this occurs in $\frac{n(n-1)\ (n-2)}{1.2.3}$ ways (because the number of combinations that can be made of n things taken r at a time is $\frac{n(n-1)\ (n-r+1)}{r!}$) the probability that two will fail and $(n - 2)$ happen is $\frac{n(n-1)\ (n-2)}{1.2.3} p^{n-3}q^3$.

Thus, if $(P + Q)^n$ is expanded by binomial formula,

$$(P + Q)^n = P^n + nP^{n-1}Q + \frac{n(n-1)}{1.2} P^{n-2}Q^2 + \dots$$
$$n(n-1)\ (n-2).\ (n-r+1)\, P^{n-r}Q^r +$$

The binomial expansion. But the first term of this expression (called the binomial expansion), P^n is the probability that all events will happen; the second term of the expression is the probability that 1 will fail; the third term is the probability that 2 will fail, etc. Thus each successive term in the binomial expansion represents the probability of all events happening, all but one happening, all but two, etc., throughout the series. We thus have a general expression to aid us in determining the probable frequency of occurrence of compound events contributed to by various assignable causes. To illustrate the method we ordinarily turn to such cases as coin tossing, or dice throwing, in which the chance of an event can be definitely assigned.

If the chance of an event happening or failing is known, — as in coin tossing (*i.e.*, if we let $p = \frac{1}{2}$, $q = \frac{1}{2}$), n may be assigned any desired value and the separate constituent probabilities figured. For example, if we toss 7 coins, say 1280 times, and record the number of "heads" each time, we should get theoretically, from the binomial expansion: —

$$\left(\frac{1}{2}+\frac{1}{2}\right)^7 = \overset{7H}{\frac{1}{128}} + \overset{6H}{\frac{7}{128}} + \overset{5H}{\frac{21}{128}} + \overset{4H}{\frac{35}{128}} + \overset{3H}{\frac{35}{128}} + \overset{2H}{\frac{21}{128}} + \overset{1H}{\frac{7}{128}} + \overset{0H}{\frac{1}{128}}$$

The degree to which an actual distribution checks the theoretical expansion is shown by the following distribution of heads and tails obtained by 10 students, each tossing 7 coins 128 times.[1]

7 Heads	*6 H*	*5 H*	*4 H*	*3 H*	*2 H*	*1 H*	*0 H*
1.1	7.0	21.6	36.8	33.3	20.2	6.9	1.1

It should be noted that in the tossing of the seven coins, nothing is more uncertain than that *a particular coin* will fall heads, but experience is found to check closely the theoretical statement that, in the long run, coins will fall heads and tails in proportion to the frequencies stated by the terms of the above expression. This checks the point made above, that while we expect great fluctuation in the sizes of particular individuals selected from a total group of measures, if successive groups of considerable size are drawn out we expect constancy of average values. Recall here that we need these statements of probability because we are constantly dealing with selected samples of total groups of very

---

[1] Data from H. L. Rietz, University of Illinois.

large numbers, and are forced to make statements about them in terms of the most probable situation.

For an ideal case like coin tossing, where $P = Q = \frac{1}{2}$, the binomial expansion becomes: —

$$\left(\frac{1}{2}+\frac{1}{2}\right)^n = \left(\frac{1}{2}\right)^n + n\left(\frac{1}{2}\right)^n + \frac{n(n-1)}{1 \cdot 2}\left(\frac{1}{2}\right)^n + \frac{(n)\,(n-1)\,(n-2)}{1 \cdot 2 \cdot 3}\left(\frac{1}{2}\right)^n + \ldots\ldots$$

If $n = 4$, the probabilities are as stated on page 195.

$$\left(\frac{1}{2}+\frac{1}{2}\right)^4 = \frac{1}{16}+\frac{4}{16}+\frac{6}{16}+\frac{4}{16}+\frac{1}{16}$$

If $n = 6$, $\left(\frac{1}{2}+\frac{1}{2}\right)^6 = \frac{1}{64}+\frac{6}{64}+\frac{15}{64}+\frac{20}{64}+\frac{15}{64}+\frac{6}{64}+\frac{1}{64} = 1$

If $n = 8$,

$$\left(\frac{1}{2}+\frac{1}{2}\right)^8 = \frac{1}{256}+\frac{8}{256}+\frac{28}{256}+\frac{56}{256}+\frac{70}{256}+\frac{56}{256}+\frac{28}{256}+\frac{8}{256}+\frac{1}{256} = 1$$

Probable frequency polygons. In Diagram 33, we give a graphic representation of the distribution of the probable frequency of occurrence of various events when it is possible to assign values to p and q. The student will perceive that making p and q equal, leads to a symmetrical distribution: with an odd number of terms, there will be one middle term with ordinates distributed symmetrically on both sides; with an even number, — two middle ordinates equal in size. Making p and q equal thus results in symmetrical polygons that seem to approximate the shape of distributions that have been found to fit various human traits.

Each of the successive terms in these expansions represents the *chance* of getting a given "combination" of causes in contributing to a particular event. To make clearer what we have here, let us plot frequency polygons, as in Diagram 33. Here, the heights of the ordinates erected at equal intervals on the horizontal line (the X axis) represent, to scale, the relative probability of the various events happening. For example, in the polygon for $n = 8$, the height of the extreme ordinate 1, represents a probability, $\frac{1}{256}$, that, say 8 heads (or all of 8 *like* human causes) might occur together in one throw of 8 coins (or in one sample including

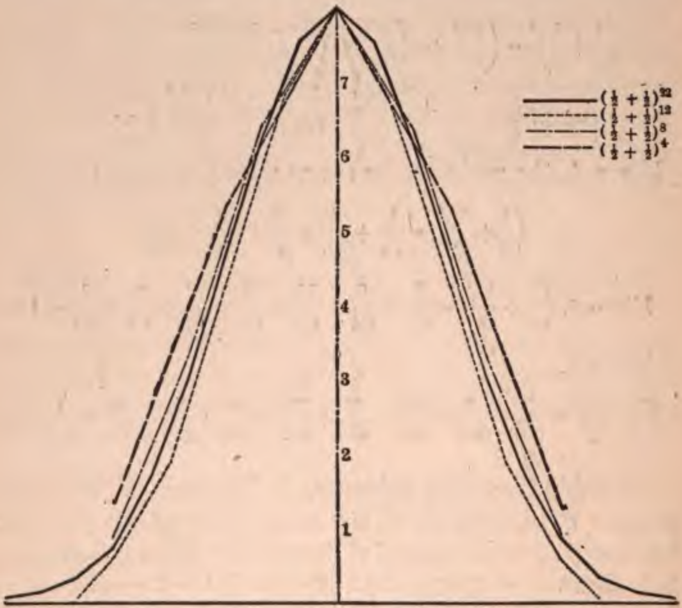


DIAGRAM 33. POLYGONS REPRESENTING THE EXPANSIONS $(\frac{1}{2}+\frac{1}{2})^4$, $(\frac{1}{2}+\frac{1}{2})^8$, $(\frac{1}{2}+\frac{1}{2})^{12}$, $(\frac{1}{2}+\frac{1}{2})^{32}$

Height of mean ordinate taken equal to 8 units; other ordinates in proportion to relative sizes of coefficients of the expansions. Abscissæ are approximated in length so as to make the polygons for different exponents similar. If the "normal curve" had been drawn the closeness of fit between it and $(\frac{1}{2}+\frac{1}{2})^{32}$ would be evident.

these characters). That is, it is probable that, in the long run, once in 256 times all 8 coins would fall heads; the second ordinate 8 indicates that it is probable that, in $\frac{8}{256}$ths of the times, 7 will fall heads and 1 tail, etc., throughout the distribution. It can be seen that, as we increase the number of independent contributing factors (n), our distribution continually approaches a smooth curve as a limit. For example, the polygon plotted from the expansion $(\frac{1}{2}+\frac{1}{2})^{32}$ is shown by Diagram 33 to approximate very closely such a "continuous" curve. It can be seen that further increase of the number of cases refines very little, for practical purposes, the apparent continuity of the distribution.

It should be noted at this point that the sum of the heights of all the ordinates in any one of our polygons represents the total number of measures. It also represents the sum of the separate probabilities, which we found must be certainty, or 1, regardless of the value of n.

The practical question[1] now arises: How use the polygons plotted from various binomial expansions to help us in interpreting our actual frequency distribution? Is it possible to compute the terms of an expansion (and thus plot the polygon) comparable to the distribution of our actual data? We can answer at once: It is possible to do this, but to do so involves both a great deal of arithmetic labor (for example, the computation of many terms of a binomial series) and methods of approximation in computation. To check the interpretation of our data against an ideal frequency curve we certainly need shorter methods than would be involved in the use of "probability polygons." We need, for example, to replace the polygon by a continuous curve which will

---

[1] For the mathematically trained student it should be pointed out that our binomial expansion is a case of discontinuous variation, and we need a method of passing from such to a curve representing continuous variation.

have ordinates approximately the same relative height, and which will be so built that the area between *any two* ordinates, say y_1 and y_2, will give the relative frequency of the measures between the two corresponding values of x, say x_1, and x_2.

The normal, or probability, curve. The continuous curve which does this is known variously as: the probability curve; the curve of error; the normal frequency curve; the Gaussian curve, or the La Place-Gaussian curve, after Gauss and La Place who separately developed the equation for it. We shall refer to the curve hereafter as the normal curve or the probability curve. The equation of the curve is developed by certain investigators in accordance with criteria obtained from the binomial polygon $(\frac{1}{2}+\frac{1}{2})^n$, and may be stated at once as: —[1]

$$y = y_o e^{\frac{-x^2}{2\sigma^2}}$$

In this equation e is the constant 2.71828, known as the base of the Napierian logarithm system. y and x are the two variables, x the distance taken on the base line of the curve from the mean to a given point, and y the height of the ordinate erected at that point. y_o and σ are two very significant terms in the equation of the curve. σ is the standard deviation of the distribution, which the student has already met in computing variability. Thus if the deviation of each measure is taken from the arithmetic mean, and called d,

$$\sigma = \sqrt{\frac{\Sigma d^2}{n}}$$

and in Chapter VI, it was pointed out that *it is a unit of distance on the scale which can be used to describe the relative*

---

[1] In order to make definite use of probability and correlation methods the student will be forced to review slightly his elementary algebra. Chapter IX gives a discussion of equations and their plotting, which may also be helpful at this point.

amount of variability of the measures around the mean. The student should stress the fact that the unit of variability, σ, which he has learned to compute numerically, *is exactly the same unit distance on the X axis*, as now enters as a measure of variability of x in the equation of the curve.

For an adequate comprehension of the graphical significance of σ the student must study the way in which the equation of the curve is built. *Note that any distance on the X axis (i.e., any "x") is* measured in units of σ. Familiarity with this is absolutely necessary.

Note furthermore that as you let x take various values, y is always expressed as a proportional part of y_o. That is, when $x = o$, $e^o = 1$, and $y = y_o$. Thus y_o is the greatest ordinate and all of the other ordinates of the curve will be expressed as fractional parts of y_o. Furthermore the curve is symmetrical about the point $x = o$, and the arithmetic mean, median, and mode coincide at this point. The term y_o may also be computed by the equation:

$$y_o = \frac{N}{\sigma\sqrt{2\pi}}$$

where N = the total number of cases, σ is the standard deviation and π is the constant 3.1416. Thus the complete equation of the curve, as written by followers of Pearson and measured in units adaptable to the data of educational research, is:

$$y = \frac{N}{\sigma\sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}}$$

This, then, *represents* the *normal* probability curve, *taken to typify, approximately enough for practical purposes, many human traits* in which educationists are primarily interested. Several practical questions must next be answered concerning the use of it to such students.

ILLUSTRATIVE PROBLEMS[1]

1. Compute the first and second "smoothed" frequencies of the data in Problem No. 4, Chapter IV (distribution of monthly salary paid to teachers of science in 147 Kansas High Schools). Plot a frequency polygon for (*a*) the original distribution; (*b*) the first smoothing; (*c*) the second smoothing. Tabulate the three sets of frequencies below the base line of each graph.

2. (Review problems on graphing).

a. Plot frequency polygons for the data of Problems No. 1, 2, and 3 in Chapter IV. Arrange these three polygons on one sheet.

b. Plot the data of above problems in *column diagrams*. Arrange the diagrams on one sheet, making them as large as possible to fit the sheet.

c. Show graphically on each of these graphs the position of the mean, the median and the mode (see computations on original problems) and represent the value of the mean deviation, the quartile deviation, and the standard deviation.

---

[1] Quoted from Rugg, H. O., *Illustrative Problems in Educational Statistics*, published by the author to accompany this text. (University of Chicago, 1917.)

CHAPTER VIII

USE OF THE NORMAL FREQUENCY CURVE IN EDUCATION

HAVING established a type or ideal frequency distribution, how may we make use of it? Four definite questions must now be answered: —

1. How is the normal curve plotted in general?

2. How may it be superimposed on any actual frequency polygon to permit of direct comparison of actual and theoretical distributions?

3. How may the normal curve be used to determine the number or proportion of the individuals that ought to fall between any two selected values; *e.g.*, in the marking problem above, how many pupils theoretically ought to get A, B, C, etc.?

4. How may the curve be used to determine the probable reliability of the statistical results obtained from actual data?

1. How to plot the normal curve

To plot or graph a curve we need the equation of the curve. Having that given, *e.g.*, $y = 4x + 8$, our problem consists of three steps: —

1. Solving y for various assigned values of x; *e.g.*:

Let $x =$	Then $y =$
1	12
2	16
3	20
4	24
etc.	etc.

2. Laying off on the axes of X and Y, corresponding values of x and y and plotting the points determined by them.

3. Connecting the points thus plotted, to give the graph of the line (*e.g.* in Diagram 34 the line "represents" the equation $y = 4x + 8$).

To plot the equation of the normal curve

$$y = y_o e^{\frac{-x^2}{2\sigma^2}}$$

evidently necessitates much more elaborate preliminary computation than is true of this simple illustrative problem. Furthermore, there are evidently two more terms, σ and y_o, in the equation that need to have values assigned to them. Since the equation implies that all ordinates to the curve (erected at distances from the mean equal to particular fractional parts of σ) are constantly proportional to a fraction of y_o, our work would be much facilitated if we had a table in which were stated values of the ordinates to the curve (y's) corresponding to assigned values of x. For example, in Table II it is noted that the ordinate erected at $.1\sigma$ from the mean $= .995$ of the height of the mean ordinate y_o; that at $.3\sigma = .956y_o$; that at $1.0\sigma = .606y_o$; that at $2.0\sigma = .135y_o$, etc. The computation of values of y, then, for cor-

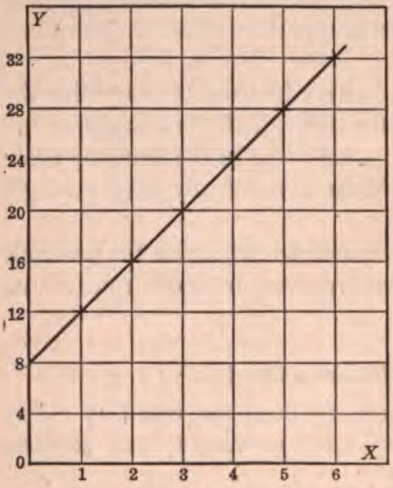


DIAGRAM 34. GRAPH OF THE LINE $y = 4x + 8$

responding values of x can evidently be done once for all and the results compiled in a table. This has been done, and Table II gives the results. *Note carefully that the computation necessitated measuring x in units of σ, and y in units of y_o.* Recall here that σ, the standard deviation of the distribution, is the fundamental unit of distance on the scale (the x-axis); also that the equation of the normal curve is so stated that y_o is the ordinate of greatest height, and that it is a constant. Therefore, the table has been derived by letting σ and y_o both equal 1, with consequent values of both x and y represented as fractional parts of σ and y_o.

Steps in plotting the curve. To plot the curve then, our steps of procedure are clear: —

1. Lay off distances on the x-axis equal to fractional parts of σ, say $.1\sigma$, $.2\sigma$, $.3\sigma$, etc., out to, say 3.0σ. *Note that the selection of the magnitude of these unit distances is entirely arbitrary, — hence that the exact shape* of the curve will depend upon the units selected.

2. Select a unit of scale for the *y's* which will give a reasonably *steep* curve, and erect at the middle point of the x-axis (*i.e.*, $x = 0$) an ordinate equal to y_o (*i.e.*, equal to 1). Note again that the *actual length of y_o* (*i.e.*, the unit of scale for the *y's*), is entirely arbitrary. Your aim should be to take such a unit on y, that, in connection with the unit on x, your final curve will be fairly steep.

3. At each of the selected points on the x-axis, $.1\sigma$, $.2\sigma$, $.3\sigma$, etc., erect an ordinate equal in length to the fractional part of y_o that is indicated in Table II. For example, for

$x = .1\sigma$ $\qquad y = .995\ y_o$
$x = .2\sigma$ $\qquad y = .980\ y_o$
$x = .3\sigma$ $\qquad y = .956\ y_o$, etc.

4. Connect the tops of the ordinates thus erected, giving the normal frequency polygon desired. The student will

note that the more closely together the ordinates are taken (*i.e.*, the smaller the fractional parts of σ), the more closely will the probability polygon approach a smooth curve.

2. How to compare an actual frequency polygon with the normal frequency curve

We have just seen that to plot a normal curve we need but two items, values of y for corresponding values of x, and that these may be computed in fractional parts of y_o and σ. In order to superimpose a normal curve on an actual frequency distribution, so as to permit comparison of the two, it is necessary to find elements common to the two distributions. Examination will show: (1) that σ is common to both, that is, that the standard deviation can be computed and compared for ANY frequency distribution. Hence we can lay off distances on the x-axis, which have been computed in fractional parts of an actually computed σ; (2) we can compute the height of y_o for our actual distribution from the formula: —

$$y_o = \frac{N}{\sigma\sqrt{2\pi}}$$

where N is the total number of measures in our distribution, π is 3.1416 and σ is the standard deviation. In addition we find: (3) that the origin of the normal curve, *i.e.*, the point from which we begin to plot measurements, is at the mean of the distribution. This is another element common to both theoretical and actual distributions, for we can compute the arithmetic mean of the actual distribution. Having y_o we can superimpose the two curves by putting the means together, making $x = 0$ at the arithmetic mean of the actual distribution. The distances on the x-scale may now be laid off by multiplying each successive fractional part of σ, say, $.1\sigma$, $.2\sigma$, $.3\sigma$, etc., by the computed value of σ in the actual distribution. Then, the length of the ordinates that are to be erected at these points may be obtained by multiplying the fractional part of y_o (read from Table II), corresponding to the selected values of $\frac{x}{\sigma}$, by the computed value of y_o.

Summary of steps necessary for the comparison of an actual frequency distribution with a normal frequency curve. To bring all the above steps clearly in mind let us list them in definite order: —

1. Plot the actual distribution, by methods already discussed in Chapter IV.
2. Set the mean point of the normal curve at the arithmetic mean of the actual distribution. Call this point $x = 0$.
3. Compute unit distances in terms of σ, that will be laid off on the x-axis by multiplying fractional parts of σ, say $.1\sigma$, $.2\sigma$, or $.01\sigma$, $.02\sigma$, etc., by the computed value of σ in the actual distribution. Note that these are to be computed either in terms of class-intervals or actual units on the scale. The two must be clearly distinguished.
4. Lay off on the x-axis, to the scale used in plotting the actual distribution, these fractional parts of σ.
5. Compute y_o from

$$y_o = \frac{N}{\sigma\sqrt{2\pi}}$$

6. At the arithmetic mean of the actual distribution erect an ordinate equal to this computed value of y_o.
7. Compute the height of the ordinates corresponding to $x = .1\sigma$, $.2\sigma$, etc., by multiplying each $\frac{y}{y_o}$ from Table II, by the above computed value of y_o.
8. Erect the ordinates at the successive points on the x-axis.
9. Connect the tops of the ordinates, giving the normal curve.

3. How to determine the closeness of fit of an actual distribution to the normal distribution

Having superimposed a normal distribution on a distribution of actual measurements, how can we determine the relative closeness of fit of the two distributions? The fundamental question involved in the discussion is this: Are the differences that have been found between the theoretical frequency (y') at any point $\frac{x}{\sigma}$, and the actual frequency (y), indications of REAL differences between the type of theoretical distribution and the theoretical normal curve? Or are they so small as naturally to be expected in the taking of samples from a very large group. Note that *the normal distribution presupposes a very large number of measures*, while the actual distribution contains but a very limited portion of the total group. The question arises: Is the sample represented by our actual distribution a "random" sample — *i.e.*, one taken by chance from the total group? If we continue to take samples of the same number of measures under similar conditions, will the samples continue to give approximately the same distribution polygons, or will they be distinctly different?

The questions indicate that the only way in which statistical methods can establish the *normality* of an actual distribution is by stating the probability that actual frequencies and theoretical frequencies at any point on the scale will differ by a given amount. They enable us to say, for example, that it is relatively likely that if we continue to take similar samples from our total group, that this particular difference will occur or will not occur repeatedly. The complete explanation of such methods cannot be taken up in such an elementary presentation as this, but the following rule-of-thumb method may be given the student interested in these problems.

Simple rule for calculation. Compute the theoretical frequency, f, at any point on the scale (this corresponds to y for that point); find the difference between this and the total frequency, N, of the whole group ($N - f$). Compute what is known as the *standard error of sampling*, or *standard deviation of simple sampling* from the expression [1]

$$\sigma_s = \sqrt{\frac{f(N-f)}{N}}$$

The rule generally given for the interpretation of the relative sizes of the actual difference and the theoretical difference, σ_s, is that if the actual difference exceeds 3 σ_s, then the actual difference did not occur as a mere fluctuation in sampling. (The reason for assigning $3\sigma_s$ as the limit will be explained in the next section.) Note that the most that can be said here is *that it is probable or improbable that a particular* difference occurred as a mere fluctuation due to taking a small sample of the total population; that if we continued to take samples of similar size that there is a certain degree of probability that a smaller or larger difference would have occurred, due to the chances set up in taking the samples. Statisticians who wish to allow as far as possible for divergences from normality as being caused by the chances set up in taking samples, imply that any difference less than approximately $3\sigma_s$ might have been due merely to such fluctuations in sampling. Any difference greater than $3\sigma_s$ is believed to show the influence of constant causes which contribute to skewness.

4. How may the normal curve be used to determine the proportion of a group of measures that theoretically ought to fall between any two selected points on the scale?

Our discussion to this point has made use *of ordinates* to

---

[1] This expression comes from $\sigma_s = Npq$ where p = probable frequency of values at the selected point, and q = their probable infrequency. Thus, —

$$p=\frac{f}{N};\ q=N-\frac{f}{N}.\quad \sigma_s=\sqrt{\frac{N[f(N-f)]}{N^2}}.\quad \sigma_s=\sqrt{\frac{f(N-f)}{N}}$$

We next wish to compare this difference, computed by σ_s with the observed difference between the actual frequency and the observed frequency. Adapted from Yule, *op. cit.*, p. 309.

the frequency curves or polygons, taking them to represent the actual number or proportion of measures distributed at different points on the scale. These ordinates erected at the mid-point of the class-interval are assumed to typify all the measures in the given class-interval. It should be clearly understood, as pointed out in Chapter IV, that the measures in a class-interval are *completely* represented only by the area between the curve, the base line, and the ordinates erected at the limits of the interval. In the last chapter the point was made that we pass from the discontinuous frequency polygon represented by $(p+q)^n$ to the continuous curve represented by $y_oe^{\frac{-x^2}{2\sigma^2}}$ in order to deal with the ideal case of large numbers of measures studied under refined conditions of measurement.

Just as the sum of the length of the ordinates found between selected points on the scale represents *approximately* the number of measures, so the area between the curve, the base line, and the ordinates represents *accurately* the number of measures. The actual computation of the various portions of the area under the normal curve, between various ordinates, would be too laborious a process for use in the working of every statistical problem. Provided the area is measured in units of σ and in fractional parts of unity (because the area of a probability curve is 1), this can be done for a very large number of intervals on the x-axis and compiled in a table available for hand-book use once for all. This has been done by the refined methods of the integral calculus and given in the Appendix in Table III. The figures of the table merely state the fractional part of the total area that is found between ordinates erected at various distances from the mean. As with the construction of a probability curve by the methods described in Section 1, so here, the

x-axis is measured in units of the standard deviation, σ, and the area is given in each case for that portion of half the curve between the mean and the assigned value of $\frac{x}{\sigma}$. A few illustrative examples will make clear the construction and use of the table.

The table should be read thus: Between the mean (taken to be the origin of the measurements) and a distance, say, $1 \times \sigma$, will be included 3413/10,000th's of the entire area of the curve, or 34.13 per cent. The curve being symmetrical, this is true for $\pm\sigma$. Therefore, between the mean and $\pm\sigma$ will be included 68.26 per cent of the entire group of measures. This is the foundation for saying in Chapter VI that the standard deviation is a unit on the scale such that if it were laid off each way from the mean, ordinates erected at these points would include about two thirds of the cases. Or, to take other examples: between the mean and $.49\sigma$ is included 18.79 per cent of the measures; between the mean and $\pm 2.5\sigma$ is included 98.76 per cent of the cases. Between the mean and $\pm 3.0\sigma$, 99.73 per cent. It can be seen then that to go beyond 3.0σ from the mean adds relatively little to the proportion of the area taken. Conversely, in practical work, it is sufficiently accurate to say that $\pm 3\sigma$ includes all of the cases, — only .27 per cent being neglected. In fact, this explains our approximate rule, made in Chapter VI, that the range is about six times the standard deviation. The student will note that 2.5σ neglects but 1.24 per cent of the measures, and some kinds of practical use of the normal curve permit this.

While the table states the proportion of measures between the mean ordinate and an ordinate at any point on the scale, it can be used to compute the proportion between two ordinates erected at *any two points* on the scale. For example, if we desired to know what proportion of our group fell between

0.6σ and 1.8σ, the arithmetic is straight-forward subtraction, as follows:

$$\begin{aligned} &\text{Between the mean and } 1.8\sigma = 46.41\% \\ &\text{Between the mean and } 0.6\sigma = \underline{22.57\%} \\ \therefore\ &\text{Between } 0.6\sigma \text{ and } 1.8\sigma \quad = 23.84\% \end{aligned}$$

To illustrate the use of the curve in practical school work let us give a few concrete cases.

Use of the normal curve: (*a*) In distributing the marks of pupils' achievement in school. Refer, here, to the case of the superintendent who wished to distribute his pupils' marks somewhat in accordance with the probability curve. Three assumptions are necessary: *first*, that concerning the number of groups into which his marks shall be thrown; *second*, that intellectual ability in the high school accorded fairly well with the probability curve;[1] *third*, that it is "practically" justifiable to break off the base line of the curve at 2.5σ.

Assuming that he wished to group his marks in five groups, A, B, C, D, E, each group representing equal intervals of ability, how many individuals should get A, B, C, D, and E respectively? First: 5 groups A, B, C, D, E, are to be distributed equally over the *entire scale*. A serious question arises here because the curve does not meet the x-axis except at infinity. However, it approaches it very closely somewhere about 2.5σ to 3.0σ. At what distance shall we break off the curve? This is evidently a matter to be determined by trial, in order to determine which length of base line divided into fifths, gives the best practical working scale with which to measure intellectual ability. Let us try cutting it off at both 2.5σ and 3.0σ, comparing the relative frequencies with both scales.

Diagrams 35 and 36 give the data for both methods of

---

[1] There is no attempt in this textbook to justify such an assumption.

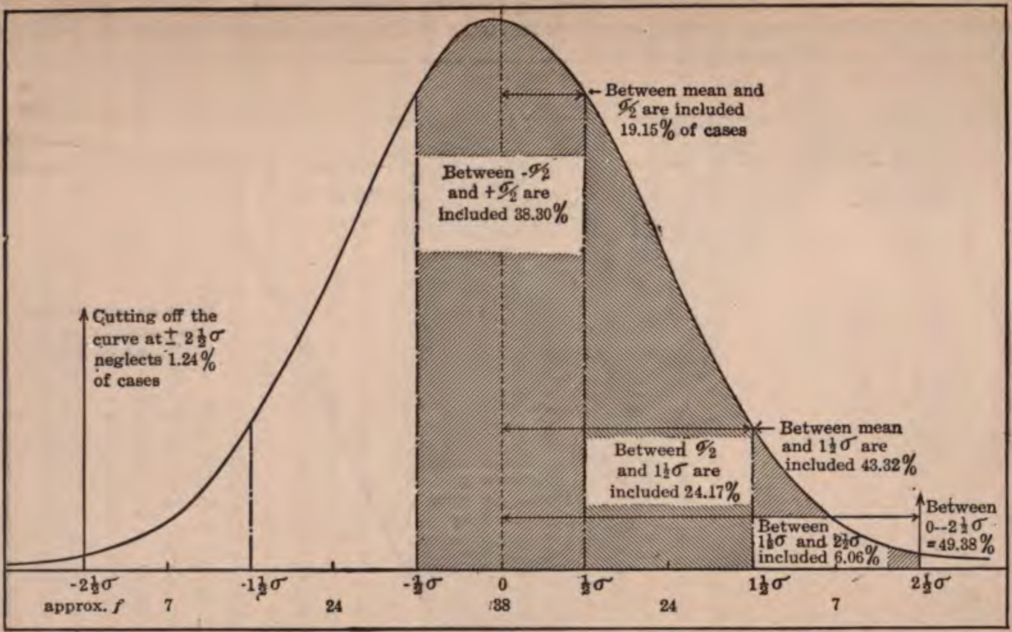


DIAGRAM 35. DISTRIBUTION OF MEASURES IN FIVE GROUPS UNDER THE NORMAL CURVE, WHEN UNIT LENGTH FOR EACH GROUP EQUALS 1.0σ; BASE LINE EQUALS 5σ

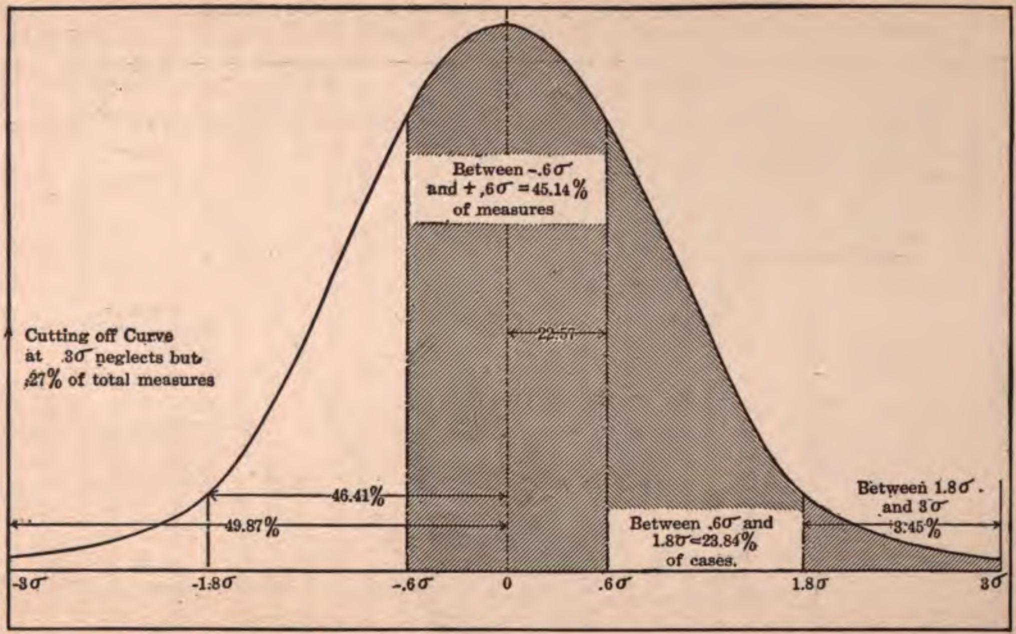


Diagram 36. Distribution of Measures in Five Groups under Normal Curve when Unit Length of Each Group equals 1.2σ; Base Line equals 6σ

fitting the normal curve to the marking system. Breaking off the curve at 2.5σ and 3.0σ respectively neglects 1.24 per cent and 0.27 per cent of the measures respectively, and gives the following distribution of measures: —

	Percentage of measures assigned to each group				
	A	*B*	*C*	*D*	*E*
Range = $\pm 2.5\sigma$	7	24	38	24	7
Range = $\pm 3.0\sigma$	3.5	24	45	24	3.5

Insufficient comparison of actual distribution of intellectual ability objectively determined, has been made as yet to permit us to make final judgments as to which of these methods gives the truer picture of human abilities. There needs to be much careful objective testing of mental functions, with very intensive comparison of actual with various theoretical distributions. The writer has been making such analyses of collected data. For practical purposes he regards a five-fold distribution with range equal to $\pm$ 2.5σ as reasonably representative. This range has been taken by other workers in the field, notably by Ayres in the design of his *Scale for Measuring Spelling Ability*. It should be pointed out that Ayres was the first to make this particular kind of practical use of the normal curve in education, in his handwriting scale as well as in his spelling scale.

Use of the normal curve: (*b*) In determining the difficulty of test questions and problems: The traditional method of designing school tests has been entirely subjective, and generally has involved the equal weighting of all questions or problems on the test; — this, too, in spite of the fact that problems and questions vary very widely in difficulty. The aims of the recent testing movement have included the determination, as closely as possible, of the real difficulty of problems used on Standard Tests. The principles and methods of procedure are illustrated in a recent article by the writer,[1] on the design of test problems in first-year algebra, an extract from which will be quoted at this place.

Principles of design of "verbal" tests in first-year algebra. "In designing tests containing verbal problems: we should (*a*), design tests of verbal problems ranging in degree of difficulty from very easy problems (which nearly all pupils will solve correctly) to very difficult problems (which but few pupils will solve correctly); (*b*), *weight each problem* in scoring the ability of pupils *by determining its relative degree of difficulty;* (3) this can be done by (*a*), finding the percentage of a large and representative group of pupils that solve each problem correctly; (*b*), assuming that algebraic abilities are distributed in the general first-year high-school population in accordance with some known distribution curve.

"Working on these hypotheses and principles, lists of verbal problems (totaling 51 in all) were drawn up covering the principal types of subject-matter named above. As a result of giving the 1915 tests, problems of widely varying degree of difficulty were included. These problems were then worked by 1295 pupils, distributed throughout 26 school systems, 17 of which also worked the 11 formal tests. As a result of this testing *there was determined the percentage of the group that worked each problem correctly*. In order then to determine the relative difficulty of each problem, the assumption was made that algebraic ability is distributed fairly closely in accordance with the "normal" probability curve. (Intellectual abilities in the elementary school have been shown to follow this distribution rather closely. We recognize the possible existence of many factors which tend to make the secondary-school curve skewed to the high end of the scale. Almost nothing is actually known of the amount and direction of their influences, however. The best "practical guess" that can be made at the present time as to the distribution of scholastic abilities is that it corresponds closely enough to the curve of error to warrant using the well-worked-out properties of that curve in our design.)

"Let Diagram 11 (Chapter I) represent the distribution of algebraic abilities in the pupils represented by our 27 school systems. The base line then represents a 'scale of algebraic difficulty' ranging, let us say, from nearly 0 ability to nearly perfect or 100 per cent

---

[1] "*Standardized Tests and the Improvement of Teaching in First-Year Algebra*"; in *School Review*, Feb. and March, 1917.

ability. The area between the curve and the base line represents the number of pupils in our entire group. If we divide the base line into any number of parts, and erect upright lines at the points representing these parts, we could determine, from the properties of the normal curve, the number of pupils that ought to be found between these distances on the base line.

"In the same way we could determine what percentage of our group of pupils should be found distributed between the zero-point on the base line and any other point. Since the normal curve has the property that it actually meets the base line only at infinity, we are forced to set our 0 and 100 points arbitrarily by deciding how large a percentage of the entire group we may drop off at both ends of the base line.

"Taking as our unit of measurement on the base line, sigma, the 'standard deviation' of the distribution (indicated graphically in Diagram 11), and laying it off 2.5 times each way from the mid-point of the curve, gives us 5 divisions (which may conveniently be divided into 10 divisions corresponding 'practically' to our public-school marking system). In doing this we are arbitrary to the extent of neglecting only 0.62 of 1 per cent of our pupils at each end of the base line. If this 0.62 of 1 per cent is thrown into the middle of the curve, where the individuals are more closely grouped, it is a negligible factor. Calling the point $2.5 \times$ sigma from the mid-point 0, setting the mean at 50, and setting the successive points 10, 20, 30, etc., to 100, at $.5\sigma$, $1.\sigma$ 1.5σ, etc., we now have a practical working 'scale of algebraic difficulty' over the successive points of which the corresponding percentages of our pupils may be indicated. Doing this, we see in Diagram 11 the proportions of our group of pupils that correspond to various degrees of difficulty on the base line. Thus a problem which is failed by 96.6 per cent of the group falls at the point marked 85; that failed by 84.8 per cent is scored 70, etc., throughout the list. To enable us to mark in an accurate way, a table has been computed in which the base line has been divided into 500 parts."

Probability table and its use. It will be noted that this application of the normal curve to school research demands a new kind of probability table, — one in which the percentage of total measures between 0 and any selected distance on the x-axis, is computed from one end of the range to the other end, instead of from the mean point, either way. In order to score problems for difficulty, which have been solved correctly by percentages of the entire group of pupils varying from nearly 0 per cent to nearly 100 per cent, we need a probability table, giving distances on the base line corresponding to various percentages of pupils. Tables V and VI in the Appendix give such data, having been constructed from Table III by setting 0 arbitrarily at -2.5σ and -3.0σ respectively, and by subtracting the total percentage between the mean and successive points on the x-axis. For example, between 0, (-2.5σ) and $.01\sigma$ is included .02 per cent of the total group; between 0 and say $.5\sigma$ is included 1.66 per cent of the total measures; between 0 and σ, 6.06 per cent, etc. Thus, if a problem is failed by only 6.06 per cent, its difficulty will be indicated by its position on the base line of the curve. These values *could* be stated in multiples and parts of σ, computing from the arbitrarily chosen end of the base line at the left. To turn the scores of the problems into 'practical' accordance with the usual percentile marking scale, each point on the base line has been transmuted into such a 100 per cent scale. To do this the mean has been set arbitrarily at 50. Since there are 10 divisions (each $.5\sigma$ in length) the even points on the scale are as indicated. Thus any problem failed by 98.34 per cent of the pupils is scored 90; that failed by 93.94 per cent is scored 80, and so on over the scale. Both Tables V and VI, in the Appendix, and Diagram 11 give the details of the method.

5. *Use of the normal curve in giving "credit for quality"*

Many school systems, at the present time, are giving additional credit to those pupils who maintain a high standard in their work:.— "credit for quality" the scheme is called. For example, in certain schools it has been customary to classify marks in five groups, say, — Excellent, Superior, Medium, Inferior, and Poor, and to weight these in such a proportion as 1.2, 1.0, 0.8, 0.6, 0.4, respectively. Such a weighting of the different grades is of course entirely arbitrary. In view of the growing interest in this practical problem a method will be suggested of weighting ability more definitely in accordance with its distribution.[1]

The method implies two assumptions: *First*, scholastic abilities distribute roughly in accordance with the normal curve; *second*, the school should give credit for achievement roughly in inverse proportion to the frequency of the pupils who reveal the achievement. In support of this latter, — the social world pays most for those things that few of its members can do, — in the same way the school should reward most highly those types of achievement that relatively few pupils can attain. Having made these two assumptions it is possible to present a workable and defensible method of crediting various grades of achievement. Since we are dealing with the probability curve, one other assumption is necessary — namely, that concerning the point at which we must break off the base line of the curve. In the light of what has been said in the foregoing sections we shall assume, *thirdly*, that human abilities are best described by a distribution (grouped, say, in five groups to accord with prevalent practice) whose range extends from -2.5σ to $+ 2.5\sigma$. Turning to Diagrams 35 and 36 we note the relative position of each of the five groups, and that *each may be typified by the median value of the class-interval* which is repre-

---

[1] The discussion of Section 5, is offered merely as a suggestion for the scientific *weighting* of student-work. The writer wishes to make it clear, however, that he is *not* a protagonist of the doctrine of giving "*credit for quality*." He regards such a practice as an administrative makeshift which would be unnecessary with the proper classification of students and courses of study. Rather, would he support the movement to segregate pupils in terms of ability with the parallel construction of a marking system which will measure abilities adequately.

sented by its portion of the base line of the curve. That is, let the entire group be represented by the values, 4.3σ, 3.4σ, etc. In the *relative distances of these mean points from the zero end of the scale* we have a definite suggestion for weighting each of the successive grades of ability. Thus, in the light of the above assumptions our method would lead to the following weights for the various grades of ability: —

	Base line of curve extends from -2.5σ *to* $+2.5\sigma$
Excellent	4.3
Superior	3.4
Medium	2.5
Inferior	1.6
Poor	.7

It can be seen that although the *absolute* values of the mean points in terms of σ are different, the *relative* distances are the same.

6. *Use of the normal curve to determine the reliability of the statistical results obtained from actual data*

Of all the uses to which we put the probability curve we now come to the most important, namely — the determination of the reliability of statistical measures, such as averages, measures of variability, and measures of relationship.

The student should recognize clearly that he always will deal with but a limited number of measures from the total group. Thus his computed average, for example, is not the "true" average, for the "true" average could be obtained only from *all* of the measures in the entire population. The "*true*" average spelling ability in the sixth grade of a large city system could be found only by testing *all* of the 20,000 children, say, in all of the sixth grades of the system. It is inexpedient, however, to test all, and so we are forced to deal with but a "sample" of the total. It has been pointed out that, to permit sound conclusions for the whole group from the statistical treatment of "samples," such samples must be "*random.*" That is, they must be, first, large enough; and *second*, chosen purely by chance, so that statistical measures (such as averages) which are computed to represent them will not fluctuate seriously in value as successive samples are chosen of the same size and in the same manner. It is evident that the weak spot in this statement is the phrase "will not fluctuate seriously." At once, we wish to know how large is a "serious" fluctuation in the size of our constant. Or, more technically, is the deviation large enough to be an indication that constant causes are contributing to our average to such an extent that our sample is not random.

To answer these questions we are forced to turn to the theory of probability. Assuming that the entire distribution from which our sample data are drawn fits the "probability" curve, and that our successive samples do so also, we can make a statement concerning the *probable deviation* of our computed constants from the corresponding "true" values. Now, we know that the most probable "error" or "deviation" from a true average, say, is the error or deviation 0. That is, if we are computing the average spelling ability of our 20,000 pupils by taking successive samples of 200 each, the average computed for each sample will be in "error" or "deviate" from the true average by some definite amount. The best assumption that we can make about the distribution of such "errors" is that they accord with the "curve of error," the probability curve. Thus in Diagram 37, we can represent the shape of the distribution by placing the mode of the probability curve at 0 error. Now, for the sake of illustration, assume that each of the 100 successive groups of 200 pupils has been tested, that the average achievement of each group has been determined, that the "true" average achievement of the whole 20,000 has been determined, and likewise the "deviation" of the average of each sample. It has been shown in various problems of measurement that these 100 deviations or errors will tend to distribute themselves in the form of a probability curve, with the mean at the error zero. Suppose, for illustration, that we assume that the actual averages computed are given in Table 32.

TABLE 32. AVERAGES FOR SPELLING ABILITY OF 20,000 SIXTH-GRADE CHILDREN (HYPOTHETICAL)

Classification of "average achievement" of samples of 200 pupils each	*Value of mid-point of each interval*	*Corresponding "error"*	*Frequency. Per cent of total number of samples which show particular "errors" in average value*
74.1–74.3	74.2	+1.0	1
73.9–74.1	74.0	+ .8	2
73.7–73.9	73.8	+ .6	3
73.5–73.7	73.6	+ .4	11
73.3–73.5	73.4	+ .2	21
73.1–73.3	73.2 = true average	0	25
72.9–73.1	73.0	− .2	21
72.7–72.9	72.8	− .4	10
72.5–72.7	72.6	− .6	3
72.3–72.5	72.4	− .8	2
72.1–72.3	72.2	−1.0	1

In this particular case, assuming that such measures of successive samples have been computed and plotted, we can express the probable deviation of the computed averages from the true average by reference to the table. Since 67 per cent of the measures fall between actual average values of 73.5 and 72.9, and 33 per cent fall outside (or, in other words, 67 per cent show a deviation of less than ±.2 per cent), the chances are 2 to 1 against a sample of 200 pupils selected at random being more than 73.5 and less than 72.9. Since 94 per cent of the cases fall between 72.6 and 73.9, or within a deviation from the true average of ± .6, the chances are roughly 16 to 1 against the average value of any sample of

100 pupils being more than 73.9 or less than 72.6. By extending our scale of error we can find a "*distance on the scale*" beyond which it is practically certain that the computed constant will not fall. That is, by reference to the theory of probability we can determine the probable extent of fluctuation of our computed constant.

7. *Various measures of unreliability*

A. STATEMENT OF UNRELIABILITY IN TERMS OF THE STANDARD DEVIATION

But, finding a "*distance on the scale*" consists in measuring variability and we have two accepted unit measures of variability, the *standard deviation* and the *probable error*. Hence we desire formulæ by which we can compute the variability of the probable deviation of computed measures from corresponding true measures. Assuming the probability curve as the form of the distribution of the deviations, formulæ have been derived mathematically for the standard deviation of an average, for the standard deviation of a standard deviation, for the standard deviation of a coefficient of correlation, and for other measures of a distribution. Note that such measures are really measures of *un*reliability.

(*a*) Unreliability of an arithmetic mean. The standard deviation of the deviation of a computed average from a true average (σ_M,) may be computed from

$$\sigma_M = \frac{\sigma_{distribution}}{\sqrt{N}} \qquad (1)$$

that is, it is equal to the standard deviation of the actual distribution of original measures divided by the square root of the number of measures. Diagram 37 illustrates its meaning.

For purposes of illustration assume

$$N = 900; \ \sigma_{distribution} = 6$$

and

$$M = 73.2.$$

Then

$$\sigma_M = \frac{6}{\sqrt{900}} = .2.$$

This means that if the *deviations* of successively computed averages from the true average of the entire distribution are

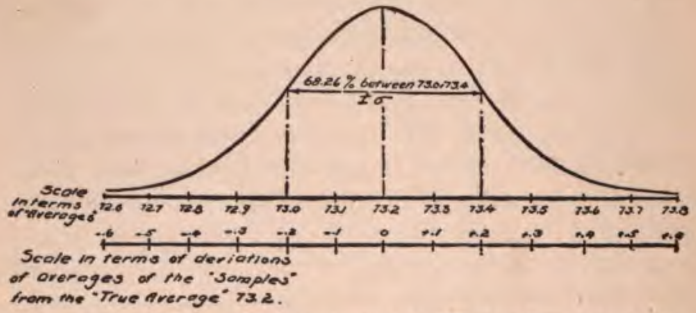


DIAGRAM 37. "NORMAL" DISTRIBUTION OF "ERRORS" IN AVERAGES

Computed for successive "samples," from "true average" of entire population. (Compare with Table 32.)

plotted as in Diagram 37, that a distance on the scale equal to σ_M will extend from $-.2$ to $+.2$. 68.26 per cent of the probable deviations theoretically will be included between $\pm\sigma_M$, *i.e.*, $\pm.2$. It was shown above that 99.73 per cent of the probable deviations will be included between $\pm 3\sigma_M$, *i.e.*, between $\pm.6$. This is interpreted to mean that the chances are about 9973 to 27, *i.e.*, about 365 to 1, that the average of any such sample selected at random will fall between $73.2 \pm .6$, *i.e.*, between 72.6 and 73.8.

(*b*) **Unreliability of a standard deviation.** In the same way we may express the probable deviation of a computed standard deviation from the true standard deviation, using the formula: —

$$\sigma_{deviation\ in\ \sigma} = \frac{\sigma_{distribution}}{\sqrt{2N}} \qquad (2)$$

The formula and method of computation may be interpreted graphically in the same way as before, remembering now that the deviations to be plotted on the scale are probable deviations of the observed σ's from the true σ.

(*c*) **Unreliabilty of a difference between two measures.** Similarly, the unreliability of a difference between two quantities may be expressed in terms of the probable deviation of the true difference from the computed difference. It can be shown that the standard deviation of this probable deviation of the difference between two measures equals the square root of the sum of the squares of the probable deviations of each true measure from its corresponding computed measure. That is, —

$$\sigma_{difference\ between\ x\ and\ y} = \sqrt{\sigma_{M\ of\ x} - \sigma_{M\ of\ y}} \qquad (3)$$

(*d*) **Unreliability of a coefficient of correlation.** It will be shown in Chapter IX that the unreliability of a coefficient of correlation is

$$\sigma_{deviation\ in\ r} = \frac{1 - r^2}{\sqrt{N}} \qquad (4)$$

The graphic interpretation will be clear to the student provided it is remembered that the scale of the base line of the curve is now, "deviations in the size of the computed correlation coefficient from the size of the true correlation coefficient."

B. STATEMENT OF UNRELIABILITY IN TERMS OF THE PROBABLE ERROR

It has been noted that there are two accepted unit measures of variability, the standard deviation (σ) and the probable error, $P.E.$ The relation between the two can be shown to be

$$P.E. = .67449\,\sigma \tag{5}$$

This relationship can be made clear by turning to Table III, which states the fractional part of the area between the mean of the normal curve and ordinates erected at distances from the mean equal to successive increments of σ. For example, between the mean and σ will fall 34.134 per cent of the entire distribution. We *define* the probable error as *that unit distance* on the scale which, if laid off one way from the mean, will determine one fourth of the cases. Therefore we can determine from the table the fractional part of σ that one will lay off from the mean to determine 2500/10,000 of the area of the curve. This proves to be .67449 σ, or approximately .6745 σ.

Because of the "common sense" meaning of the probable error (namely, that *distance* which if laid off both ways from the mean determines half the cases) it has become customary to express the unreliability of measures in terms of the probable error instead of the standard deviation. Thus the above formulæ become: —

$$P.E._{\text{arithmetic mean}} = .67449\sigma_{\frac{\text{distribution}}{\sqrt{N}}} \tag{6}$$

$$P.E._{\text{median}} = .84535\sigma_{\frac{\text{distribution}}{\sqrt{N}}} \tag{7}$$

$$P.E._{\text{standard deviation}} = .67449\sigma_{\frac{\text{distribution}}{\sqrt{2N}}} \tag{8}$$

$$P.E._{\text{coefficient of correlation}} = .67449\ \frac{1-r^2}{\sqrt{N}}. \tag{9}$$

It is convenient for the student to have in mind the following table of statements of unreliability of measures.

The chances that the true value (of the average, standard deviation, coefficient of correlation, etc.) lies within: —

± *P.E.* are 1 to 1 (50 per cent of measures fall within ± *P.E.*)
±2 *P.E.* are 4.5 to 1 (82.26 per cent of measures fall within ± 2 *P.E.*)
±3 *P.E.* are 21 to 1
±4 *P.E.* are 142 to 1
±5 *P.E.* are 1310 to 1
±6 *P.E.* are 19,200 to 1

Thus, to insure a satisfactory degree of reliability of the computed measures conservative practice insists that the coefficient be at least four times the size of the probable error.

ILLUSTRATIVE PROBLEMS[1]

1. Make three different graphs of the normal probability curve to illustrate the differences occurring in the slope of the curve as distinctly different scales are chosen for X and Y. Plot the three curves on one sheet and use your own judgment in selecting the units for X and Y.

2. For the following data, plot a frequency polygon, choosing scales on X and Y that will give as large a graph as possible and a reasonably "steep" curve. Superimpose a NORMAL CURVE on this polygon to permit comparison of the actual distribution with the theoretical distribution.

DISTRIBUTION OF STATURE FOR ADULT MALES BORN IN GREAT BRITAIN. REPORT OF ANTHROPOMETRIC COMMITTEE TO THE BRITISH ASSOCIATION, 1883, P. 256. (QUOTED BY YULE, P. 88.) CLASS-INTERVALS ARE PRESUMABLY 57.99 − 58.99, ETC.

Height (inches)	*Number of men*	*Height (inches)*	*Number of men*	*Height (inches)*	*Number of men*
57	2	64	669	71	392
58	4	65	990	72	202
59	14	66	1223	73	79
60	41	67	1329	74	32
61	83	68	1230	75	16
62	169	69	1063	76	5
63	394	70	646	77	2

Total 8585

3. Plot a normal curve on a base line extending from − 4 *P.E.* to + 4 *P.E.* Divide this base line into 5 equal parts and erect ordinates at the points of

---

[1] Quoted from Rugg, H. O., *Illustrative Problems in Educational Statistics*, published by the author to accompany this text. University of Chicago, 1917.

division. Compute the exact proportion (to 2 decimal places) of all measures that should fall within each division of the total area under the curve. On the graph, letter the points of division in units of *P.E.* and the proportion of measures in each portion of the area.

4. Prepare a "probability table" for the normal curve on a base line extending from − 4 *P.E.* to + 4 *P.E.*, in which the zero point of the table is transferred from the mean to — 4 *P.E.* State the percentage of measures that should fall between 0 (which is set at − 4 *P.E.*) and ordinates erected at successive intervals of .2 *P.E.* on the base line. This will give a table of 40 points of sub-division.

CHAPTER IX

THE MEASUREMENT OF RELATIONSHIP: CORRELATION

Practical need for measures of relationship. The previous chapters have put before us the three methods of treating a *single* distribution of educational data: (1) that of picturing its status by computing some average to represent it; (2) that of picturing its degree of concentration by computing some measure of variability or dispersion; (3) that of graphically picturing the entire distribution by plotting the frequency polygon, column diagram, or smoothed frequency curve that may be taken to represent the most probable statement of the true situation typified by our sample. It was found that if one desired to *compare* the status of two distributions he could use these methods of averages, dispersion, and frequency curves to give a complete picture of either distribution alone, or of the one compared with the other. It is probably true that most of the actual administrative problems faced by the practical school man may involve the use of these statistical methods, and only these. However, in the analytical experimental study of problems of learning and teaching, and of some administrative problems, a new type of device is demanded, — namely, some method of determining the degree of causal connection exhibited by certain traits or activities in which we are interested. The measuring of physical, mental, and social activities constantly involves the study of causation or causal connection between two or more traits in question. The massing of data in this study of causation raises the necessity for statistical methods of computing degrees of causal connection.

Suppose, for example, that we were interested in the practical problem of classifying pupils in school in terms of abilities. One of the questions for which we desire answers would be: Are "school" abilities specialized, or general? *Is it probable* that a pupil who shows a high degree of achievement in one subject of study, say mathematics, will show a high degree of achievement in another subject, say modern languages? To illustrate the problem: the data of Table 33 represent the actual school marks given a class of 23 high-school pupils in mathematics and modern languages. Each mark in the table is the average of three or more marks in the respective subject.

TABLE 33. SCHOOL MARKS GIVEN A CLASS OF 23 HIGH-SCHOOL PUPILS IN MATHEMATICS AND MODERN LANGUAGES

Pupils	*Average mark in mathematics*	*Average mark in modern languages*	*Rank in achievement in mathematics*	*Rank in achievement in modern languages*
A	50	58	23	21
B	78	88	15	7
C	96	90	2	5
D	88	85	6	10
E	85	93	8	2
F	80	57	13	22
G	94	91	3	4
H	79	84	14	11
I	86	83	7	12
J	75	80	16	14
K	83	92	10	3
L	82	81	11	13
M	71	77	20	16
N	72	59	19	20
O	92	87	4	8
P	81	89	12	6
Q	84	76	9	17
R	74	75	17	18
S	69	78	21	15
T	97	94	1	1
U	73	86	18	9
V	66	72	22	19
W	90	50	5	23

To answer our question we now have for each pupil in the class a pair of records of achievement, *i.e.*, his average mark in mathematics, and his average mark in modern languages. If, now, there were absolutely perfect correspondence, or "*correlation*" as we shall call it, in the two abilities in question, and assuming for the time being that the school marks of these pupils adequately measure their respective abilities, each pupil should occupy the same relative position in the two series of marks; — *i.e.*, the pupil first in mathematics should be first in languages, the pupil second in mathematics should be second in languages, and so on through the list. Table 34 shows this situation by giving

TABLE 34. HYPOTHETICAL MARKS GIVEN TO 23 PUPILS; PRINTED HERE TO ILLUSTRATE PERFECT "RANK" CORRELATION

Pupils	*Mark in mathematics*	*Mark in modern languages*	*Rank in achievement in mathematics*	*Rank in achievement in modern languages*
A	97	94	1	1
B	95	93	2	2
C	93	91	3	3
D	90	90	4	4
E	89	89	5	5
F	88	87	6	6
G	87	86	7	7
H	85	85	8	8
I	84	84	9	9
J	82	82	10	10
K	80	79	11	11
L	79	78	12	12
M	76	76	13	13
N	75	74	14	14
O	73	73	15	15
P	72	72	16	16
Q	71	70	17	17
R	70	69	18	18
S	67	67	19	19
T	66	65	20	20
U	65	60	21	21
V	64	55	22	22
W	63	50	23	23

two hypothetical series of marks. *This method* of measuring the degree of correspondence between two traits obviously *takes account only of the position of the various measures in the series. It neglects the absolute amounts of the measures.*

Not only should the *position* be the same for each pupil in the two series, but, in order that the correspondence be absolutely perfect, the actual proportional differences between each two consecutive marks ought to be the same. It is clear that merely to rank the measures in the two series in order of size and to compare the corresponding ranks does not accurately measure the degree of correspondence; *i.e., it does not take full account of the absolute value of each measure.*

Need of devices to show correspondence. For this reason we need devices for picturing the correspondence between the actual measures which will take full account of the actual amount of each one. For example, let us take the pairs of marks in Table 33. In Chapter IV it was pointed out that a distribution can be completely represented by graphic methods, — by *plotting the data.* Let us plot the data of Table 33. By what graphic methods can we now combine *pairs of measures* in the same diagram to show the correspondence between two varying traits? Recall here that in the preceding chapters we have been plotting *single* distributions by laying off the *units of scale* on the horizontal (X) axis, and the corresponding *numbers of measures* on the vertical (Y) axis. In the plotting of the single distribution, therefore, we deal with but two quantities, — the value of magnitude of the measures, and the frequency with which each occurred. We now have two frequency distributions, each having *a scale* along which the measures are distributed, and a set of frequencies. It is possible to combine the two distributions, however, on two coördinate axes because they have one element in common

— the frequency column. If, now, we construct a double-entry table, like that in Diagram 38, in which the x-axis represents, let us say, the scale of abilities in mathematics and the y-axis, the scale of abilities in modern languages, it is possible to represent on this squared table every pair of

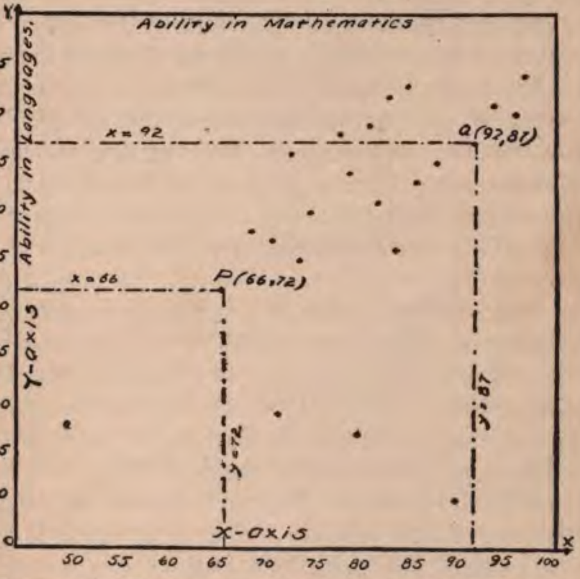


DIAGRAM 38. DISTRIBUTION OF CORRELATED ABILITIES IN LANGUAGES (y) AND IN MATHEMATICS (x)

Data of Table 33. Each point is plotted to scale to represent a pair of measures on one pupil.

measures in Table 33. Furthermore, as we do this we can see that each measure is represented in accordance with both its absolute amount and position in the series. To illustrate:

How show correlated abilities graphically. Standard usage in plotting pairs of measures involves two coördinate axes, one horizontal (OX), and the other vertical (OY), meeting at an "origin" or *beginning* point at the bottom and left. One of these axes, say OX, is chosen on which to lay off the scale of one of the traits in question, and the other axis to lay off the scale of the other trait. The selection of which trait to plot on a particular axis is left to the arbitrary choice of the student. The units of the scale are now laid off from the origin *to the right* on OX, and *upward on* OY.[1] It is now possible to represent *to scale* a pair of measures, by plotting the value of one on the x-axis, and that of the other on the y-axis. Erecting perpendiculars to the x and y axes gives us a point as the intersection. *When considered with respect to the distance which it is from either base line, $O\,X$ and $O\,Y$, this point represents the pair of measures in question.*

For example, in Diagram 38 a point, determined by perpendiculars erected at distances 50 units from the origin on OX and 58 units from the origin on $O\,Y$, *represents* the pair of marks given pupil A in Table 33, 50 in mathematics and 58 in modern languages. Similarly with pupil B, represented by a point 78 units to the right of OY and 58 units above OX; and pupils C, D, etc. It should be noted, that in Diagram 38, although the scaling of distances is correct, the entire table down and over to the origin is not given. Theoretically, of course, each point on the table is referred to axes OX and OY, assumed to be at zero.

Diagram 38 now becomes clear to us. Each point on the diagram represents a pair of measures on a pupil. All the points, considered together, typify the degree of correpondence of correlation between the two abilities. A glance at the table tells us these things: (1) in general, pupils who

---

[1] This method of plotting is in contrast to those of many educational workers in statistical methods, but more consistent with standard algebraic practice.

stand high or low in one ability stand high or low in the other; (2) there are three pupils in the class for whom very low achievements in languages accompany high or moderate achievements in mathematics. For the remaining 20 pupils the correspondence is rather close. In clarifying the situation for the investigator, however, the actual plotting of the table indicates at once, and in a much more definite way than does the ranking of Table 33, the absolute amount and relative position of each pair of measures. This method of treating the data points out that we are primarily interested in *changes* in the size of one *variable corresponding to changes in the size of the other.*

If we plot the data of Table 34 (rank correlation perfect) we have a distribution of pairs of measures as in Diagram 39. As we glance over the rank order of these 23 measures we note *perfect correspondence in change of position* of the pairs of measures in the two series; *i.e.*, pupil N is 14th in both series, pupil A is first in both, W is last in both, M is 13th in both, etc. Diagram 39, though, gives this information concerning change in position and, *in addition, shows the changes in magnitude of the various pairs of measures.* It is noted, for example, that the four smaller measures beginning with 66, 65; 65, 60; 64, 55; and 63, 50; show a much smaller decrease in the size of the x-variable (*i.e.*, achievement in mathematics) than in the size of the y-variable (achievement in modern languages). This type of eccentricity in distributions, which would not be revealed by mere ranking methods, shows up clearly in the complete plotting of the table.

Few- and many-pair correlations. Changes in the distribution of measures in a correlation-table which contains but relatively few pairs of measures, for example, 23 as above, can be comprehended rather easily. It is probable that a fairly adequate interpretation could be made of the general change in magnitude of these two variables, and expressed in word form. However, the expression certainly would be vague and consist in statements something like the following: "Large achievements in mathematics seem to be accompanied by large achievements in modern languages.

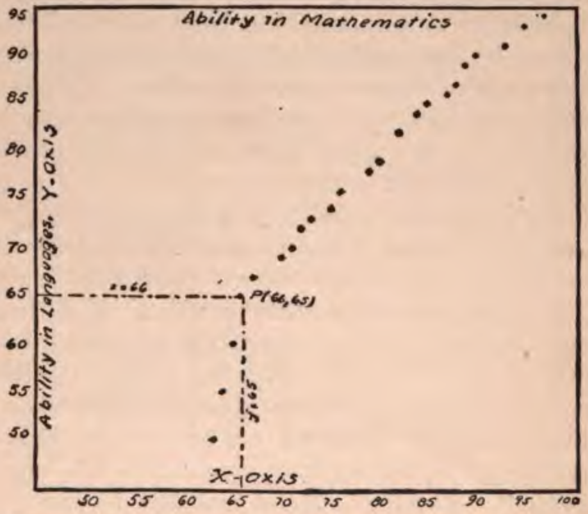


DIAGRAM 39. DISTRIBUTION OF CORRELATED ABILITIES IN LANGUAGES (y) AND MATHEMATICS (x)

Data of Table 34. Each point is plotted to scale to represent a pair of measures on one pupil.

There are certain exceptional cases, however, in which the opposite is true," etc. Most of our distributions, however, contain many measures, several hundred or several thousand in some cases. It is evident that we cannot deal adequately with the separate pairs of measures which are plotted in Diagram 40. Furthermore, as we increase the number of measures in the table the scattering of a few pairs of measures away from the mass has less and less effect on our interpretation of the general situation.

Grouping of correlated data. Now in discussing the treatment of the frequency distribution it was pointed out that the single measures may be grouped in class-intervals. In order *to condense two distributions* which have been plotted on the two axes of a double-entry table we resort to the same procedure — we group the measures on each axis in class-intervals. Doing that with a table like that

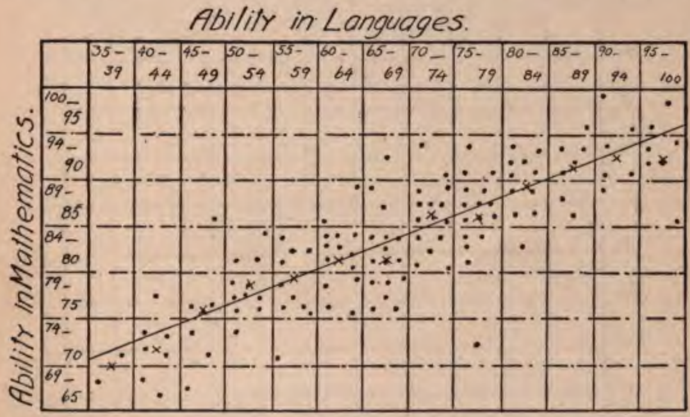


DIAGRAM 40. DISTRIBUTION OF CORRELATED ABILITIES IN LANGUAGES AND IN MATHEMATICS FOR 130 COLLEGE STUDENTS

Crosses represent mean values of points in each column.

represented by Diagram 38, we obtain a classified table like Diagram 41. In *grouping* in class-intervals, however, we must recall that the assumption is made that all measures in each square (representing an interval on each axis) are assumed to be grouped at the mean point of the square. This point is determined as the point common to the means of both axes, x and y, of the square. In Diagram 41 the measures of Diagram 38 are shown in their new grouping, each point having been moved to a position at the mean of the class-intervals. It will be noted by the student that the general shape of the distribution of the table is approximately the same. In this particular case the material is somewhat more compact, the extreme points having been moved more closely together.

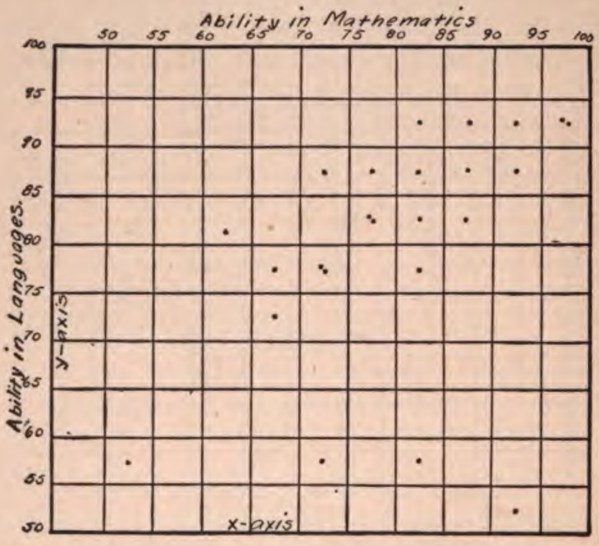
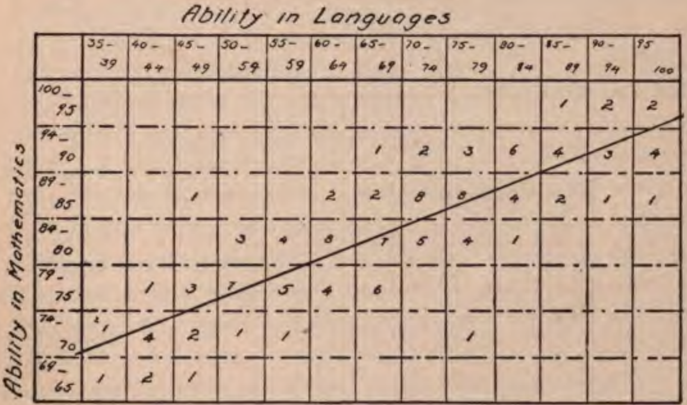


Diagram 41. Data of Diagram 38 plotted under the Assumption that all Points are concentrated at the Mean Points of the "Compartments" or Class-Intervals of the Table

This diagram illustrates "grouping" of original data in class-intervals.

To show relationship between traits. Having grouped the data in class-intervals, our next step is to tabulate in each square the number of points found to fall in that particular square. The correlation-table (Diagram 40), now becomes Diagram 42. The number in each square now represents the number of persons whose records in the two subjects fall in that particular class-interval: For example, 7 pupils received marks between 50–54.9 in languages and 75–79.9 in mathematics. Again, inspection of such a table enables general statements to be made concerning the degree of relationship between the two traits. From the general trend of the table it is evident that abilities in mathematics are *directly* related to abilities in languages.



Ability in Languages

Ability in Mathematics

	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95 100
100-95											1	2	2
94-90							1	2	3	6	4	3	4
89-85			1			2	2	8	8	4	2	1	1
84-80				3	4	8	7	5	4	1			
79-75		1	3	7	5	4	6						
74-70	1	4	2	1	1				1				
69-65	1	2	1										

DIAGRAM 42. DATA OF DIAGRAM 40 TABULATED UNDER THE ASSUMPTION THAT ALL MEASURES ARE CONCENTRATED AT THE MEAN POINTS OF COMPARTMENTS

To illustrate second step in the computation of correlated and regression coefficients.

Discovering laws of relationship. Inspection of a correlation table is not sufficient to tell us in a definite way, however, *to what degree the two are related.* If for example we have *two* correlation tables, rather similar in "scatter," it is difficult to determine by inspection of the table in which case the correlation is the more perfect. Facing such a table upon which large numbers of measures are scattered we at once feel the need for some *device for condensing the measures,* — the need of devising *an average or typical measure which will adequately represent the status of all the pairs of records* *taken together.* Just as averages and measures of dispersion typify a single distribution, so we wish a device which will succinctly and yet most completely describe the whole correlation table.

This necessary device can be constructed by turning to the columns and rows of the table. Each column or row (either of which may be called an "*array*") may be regarded as a separate frequency distribution, and as such may be typified by an average point on its scale. Remembering that the most probable value of a series of measures is the arithmetic mean of the series, we may take the arithmetic mean of each column to typify it. Doing this for each column, as in Diagram 40, we now have a fairly continuous series of mean points as we move up the table. Careful inspection of the table will show the student that these points distribute themselves in close accordance with a straight line. Thus, *in the line that will best fit these mean points we have a device for representing the entire table. The line of the means of the columns or of the rows* may be shown to represent *the most probable law of relationship* exhibited by the two variables. Nothing is of more importance to the student in studying this problem than the clear recognition of this point. "Law of relationship" implies regularity of change in the two traits,—as one grows larger or smaller the other grows larger or smaller, or *vice versa.* This may be typified by the line that most closely approximates the general scattering of the pairs of measures over the table. Now if a line can be drawn on the correlation table that will best describe or typify the law of relationship, our task is to find simple methods of dealing with such a line.

METHODS OF DETERMINING RELATIONSHIP

A. METHODS WHICH TAKE FULL ACCOUNT OF THE VALUE AND POSITION OF EVERY MEASURE IN THE SERIES

I. The Case of Straight-Line Relationship

1. The first method of determining relationship

Galton's graphic method. One's first tendency would be to deal with the graphic representation of the law — the line of relationship. That is what Galton did in his pioneer and suggestive study thirty years ago.

In Diagram 43, drawn for the data of Diagram 42 (ability in languages and mathematics), the scales on the x and y axes have been so taken that Q_3-Q_1 (for ability in languages) represents the same distance on x-axis as Q_3-Q_1 (for mathematics) on the y-axis. The points Q_3 and Q_1 have been plotted by erecting perpendiculars to OX and OY from the respective Q_3's and Q_1's on X and on Y. The heavy lines in the diagram represent coördinates drawn through the medians of both distributions. Their intersection is the median of the table. Under these conditions, and since the units of the scales on the two axes are the same, the line drawn through Q_3, Q_1 is the line of perfect correlation. In this case, it is at 45° to the horizontal base line.

Galton next drew a line to approximate as closely as possible the mean points of the actual pairs of measures in the columns (shown by the crosses). This line is seen to deviate from the line of perfect correlation. Then in the figure *any* horizontal line AB, is drawn from the median line, cutting the two lines $Q_1 Q_3$ and DB. The ratio $\frac{AB}{AC}$ measures the amount of correspondence in change in the two variables. For every point on the line of *perfect correlation*, a given change in the size of y is accompanied by a proportional change in the size of x. For every point on the line which best fits the means of the "arrays" a given change in the size of y is accompanied by a somewhat larger change in

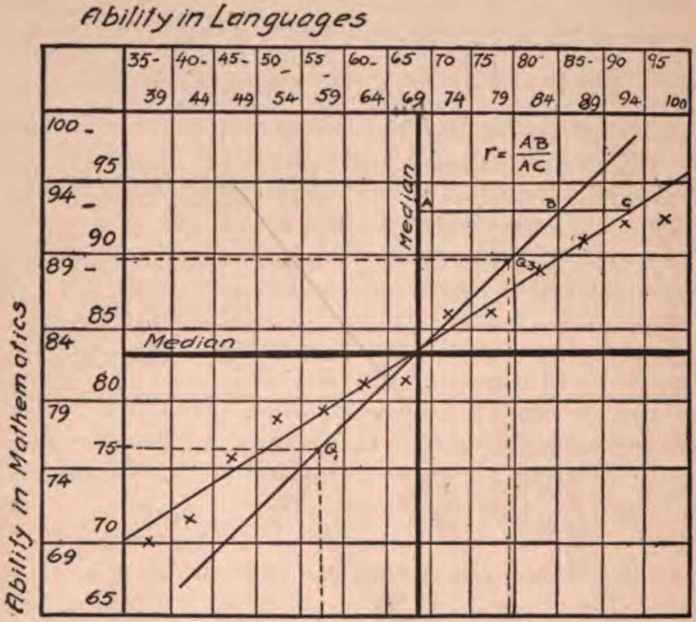


DIAGRAM 43. A GALTON DIAGRAM FOR REPRESENTING CORRELATION GRAPHICALLY

Data of Diagram 42. Scales on X and Y such that $Q_3 - Q_1$ on Y equal same distance as $Q_3 - Q_1$ on X. Thus line $BQ_3\,Q_1$, is line of perfect correlation, at 45° to horizontal.

the size of x. It is clear that if the two lines in the diagram coincide, then $\frac{AB}{AC}$ equals 1 and the correlation may be said to be perfect and positive. If the line of the means is vertical, coinciding with the median, then $\frac{AB}{AC}$ becomes 0; *i.e.*, a given change in the size of y is accompanied by no change in the size of x. As the line of the means swings over to the left of the vertical median, and its direction becomes downward from left to right, the correlation evidently becomes negative. That is, a given increase in the size of y results in a decrease in the size of x and *vice versa*. Finally when the line falls at right angles to Q_1Q_3 $\frac{AB}{AC}$ becomes -1 and correlation is perfect, but negative.

Thus Galton's method enables us to measure graphically the degree of "*co-relation*," or correspondence between two traits. Galton applied his method to the measurement of inheritance of stature by computing the coefficient of "co-relation" between the stature of children and the stature of their parents (the stature of the two parents being averaged in each case to give the "mid-parent"). He found this coefficient (the ratio described in the foregoing paragraphs) to be $\frac{2}{3}$. This may be interpreted to mean that if the average stature of a group of parents is found to be, say y inches *above or below the general average of the race*, the average stature of their children will be only $\frac{2}{3}\,y$ inches above or below the mean of the race. Galton expressed this by saying that the mean heights of offspring tended to "*regress* back toward the mean of the race." Since his time other workers in biological statistics have used his term "*regression*," and now it is common to speak of *the line of the means* of the correlation table *as the line of regression*. The ratio described above has come to be called the *coefficient-of-correlation*, and is denoted by r.

2. Second method of determining the law of relationship

Finding the equation of a straight line of regression. Refined comparative work in statistics demands a more accurate method of determining the law of relationship exhibited by two traits than that of graphic measurement. We said above that the law of relationship is described by the "*best-fitting*" or "*most representative*" line of the table, *i.e.*, by the line which fits most closely the mean points of the columns of the "arrays." Now, the most definite way by which we can describe a line is to write its equation. Since most of our educational investigations give tables whose means approximate closely a straight line, we shall confine our discussion for the time being to that type.

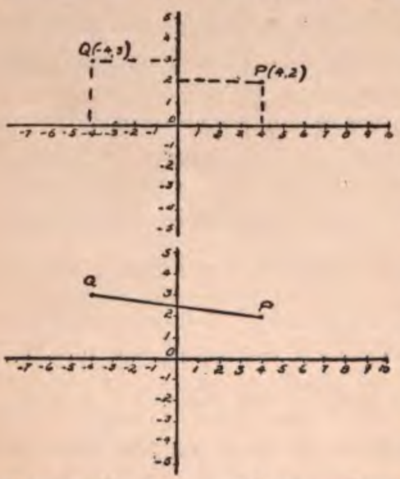


DIAGRAM 44. PAIRS OF MEASUREMENTS PLOTTED

In order to write its equation we must be able to put two variable quantities, say X and Y, together in an algebraic expression in such a way that a given change in the value of one, say x, is accompanied *by a proportional* change in the value of the other, y. For example, in Diagram 44, a series of points are plotted, each of which represents a pair of measurements, and each of which "*satisfies*" the equation of the line. That is, point P "represents," or is plotted from $x = +4$, $y = +2$; point Q represents $x = -4$, $y = +3$. The line

PQ could be completely described, therefore, by stating the pairs of coördinates of any two of its points, say $P\ (4, 2)$; and $Q(-4, 3)$. It will be noted that plotting a line, just like plotting separate points, consists in referring each point on it to the axes from which it is plotted. One characteristic of the straight line stands out, however, — all the points on this line have the property that the ratio of their coördinates, $\frac{y}{x}$, is always the same, regardless of the location of the point on the line. This ratio, $\frac{y}{x}$, is the tangent of the angle that the line makes with the horizontal axis. Since it measures the inclination of the line it is called the "*slope*," and is denoted by "m."

With this knowledge we can now write the equation of any line by making, since $m = \frac{y}{x}$, $y = mx$. m is called a "factor of proportionality." In the line plotted in Diagram 34, m is evidently 4. Thus, the equation of this line is,

$$y = 4x + b$$

and the diagram shows that taking any value for x, and computing the corresponding value for y gives a series of points, all of which fall upon this same straight line. The general slope form of the equation of a straight line is

$$y = mx + b$$

in which b is the ordinate of the point of intersection of the line and the axis of Y. In Diagram 34 it is 8.

Finding the equation for a correlation group. If we desire, now, to develop an equation or a coefficient which will describe adequately a "scatter" diagram, or correlation-table, it must be an equation which will measure in some way the *deviation of every point on the table from the means of the* *corresponding rows and columns.* The significant point for the student to master is this: — *closeness of correlation may be measured in terms of the relative amount of deviation of each point from the mean of the column and from the mean of the row in which it falls.* A table in which the measures show a high degree of correlation will be one in which the points are closely concentrated around the line of the means, — the deviations are small, as in Diagram 39. A table which shows a lower degree of correlation will reveal the measures as being very much scattered away from the line of the means, — the deviations are relatively large, as in Diagram 38. In discussing the measurement of dispersion, however, we found that, for the deviations of measures in two distributions to be comparable, they must be measured in terms of some unit deviation. The accepted unit of deviation we found to be the standard deviation. Hence our algebraic expression must *measure deviations* from the means *in units of the* respective *standard deviations.*

Now, clearly, the line which "best fits" the means of the "arrays" is that line for which the deviations of the means are the least possible. From the standpoint of convenience an equation may be derived for this line by assuming the criterion from "least squares," that the sum of the squares of the deviations of the means, each weighted by the number of measures in the respective array, shall be a minimum.

This is exactly what has been done by Professor Karl Pearson who has derived the equation of this best-fitting line. The fundamental conceptions underlying the method, however, are Bravais's, who in 1846 suggested that the correspondence of two quantities could be represented *in terms of the product-sum of the deviations from the respective means.* No single coefficient or equation was established at that time, to represent the degree of the correspondence. In 1896, Pearson published his product-moment method of computing correlation, and gave us the equation of the line of regression and the coefficient of correlation.

Pearson's equation. Working on the criterion named above, Pearson deduced the equation for the "best fitting" line as:

$$y_1 - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x_1 - \bar{x}) \qquad (1)$$

in which $\bar{y}$ and $\bar{x}$ are the mean values of the columns and rows respectively; σ_y and σ_x are the standard deviations of the two distributions, and r *is the very important statistical device known as the coefficient of correlation.*

The equation is variously given in *two* forms, one that stated in equation (1), and the other that stated as follows:

$$y = r\frac{\sigma_y}{\sigma_x}x \qquad (2)$$

in which *y and x are now deviations* of particular y and x measures from the means of the respective arrays. In other words $y = y_1 - \bar{y}$, and $x = x_1 - \bar{x}$. The student should familiarize himself with these two equations, as they are of the first importance. The theory on which they are built has led to a description of a line that safely may be regarded as *the most probable statement of the "law" represented by the data.*

Now, we note a new term in these equations — r, the correlation coefficient. We noted in the preceding section that r was the ratio $\frac{AB}{AC}$ in Diagram 43. Furthermore we said above that if the law represented by the data were typified by the "best-fitting line," the equation of the line must take account definitely of the corresponding x and y deviation of each point on the table. The process of deriving the equation of the line led to this more detailed statement of the equation: —

$$y = \frac{\Sigma[(x_1 - \bar{x})\ (y_1 - \bar{y})]}{N\sigma_x^2}(x_1 - \bar{x})$$

that is

$$\frac{\Sigma\,[(x_1 - \bar{x})\ (y_1 - \bar{y})]}{N\,\sigma_x^2} = m,$$

the "slope" of the line. Now to simplify the final statement of the equation *let us define* r as

$$\frac{\Sigma\ (x_1 - \bar{x})\ (y_1 - \bar{y})}{N\ \sigma_x\,\sigma_y},$$

or, in terms of y and x as *deviations*, —

$$r = \frac{\Sigma\ xy}{N\ \sigma_x\ \sigma_y}.$$

Then the slope,

$$m = r\frac{\sigma_y}{\sigma_x}$$

and the final equation of the best-fitting line is: —

$$y_1 - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x_1 - \bar{x}) \text{ or } y = r\,\frac{\sigma_y}{\sigma_x}x$$

(*a*) The significance of r — the coefficient of correlation. This brief mathematical statement has been given to permit us to make clear the real significance of r, the so-called "coefficient of correlation." r serves two specific functions in the determination of relationship.

(*1*) *It is a single index*, a pure number, which measures the degree of "scatter" or of concentration of the data, by giving the *mean product of the deviations of each of the measures from the mean value of its "array," when measured in units of the standard deviation.* Stated in terms of such *deviations* (x and y) r is more simply expressed as: —

$$r = \frac{\Sigma\ xy}{N\ \sigma_x\ \sigma}$$

We note that the deviation (x, or y) of each measure from its respective mean is measured in units of its respective standard deviation by dividing it by σ_x or σ_y. Σxy is evidently Bravais's product-sum of the deviations. Thus in this single numerical coefficient, r, we express relationship *in terms of the mean values of the two traits by measuring the amount each individual deviates from its respective mean.* The formula for r is generally called the *product-moment* formula. This will be explained later by reference to Diagram 46.

(*2*) *r is an intermediate device.* In defining the second function of r, we note that it is merely an intermediate numerical device, defined as it is for the purpose of bringing together, in one convenient expression, certain terms collected in the process of developing the law of regression. Thus, *it is really only an intermediate expression in the ultimate mathematical process of expressing the law of relationship in terms of the equation of the line of regression.* On the other hand, it may have for the lay student a more definite connotation than the equational or algebraic expression for the line itself which really represents the relationship; *e.g.*, —

$$y = r \frac{\sigma_y}{\sigma_x} x$$

To the mathematician this expression has a specific connotation; to the non-mathematical student a vague and unsatisfying one. Largely for that reason, students of educational research have neglected the equational expression for relationship, and have adopted the single numerical coefficient r. It should be noted, however, that once having determined r, the regression equation of the line of the means can be expressed very simply by substituting the values of r, σ_x and σ_y in the equation above.

Furthermore, we said that such an equation was expressed in the "*slope*" form, $y = mx$ in which m is the "slope" or tangent of the angle that the line makes with the horizontal. Thus in the regression equation, —

$$m = r\frac{\sigma_y}{\sigma_x}$$

and this, in those cases in which the variability of the two traits is the same, becomes equal to r. $r\,\frac{\sigma_y}{\sigma_x}$ is known as the *regression coefficient of y on x*, that is, the deviation of y corresponding on the average to a unit change in the type of x, and is represented by b_{yx} or b_1. In the same way $r\,\frac{\sigma_x}{\sigma_y}$ is *the regression coefficient of x on y*, is represented by b_{xy} or b_2, and means that deviation of x which corresponds to a unit change in the type of y.

(*b*) What is the meaning of the coefficient of correlation and the regression coefficients? Statistical measures are computed only for the purpose of clarifying our interpretation of complex masses of data. It has been pointed out repeatedly in the foregoing chapters that such devices do not supply *proofs* of existing relationships, — rather that they are merely tools to refine our analysis of numerical situations, and that they are valuable only in so far as they agree with sound logical analysis. So it is with statistical devices for measuring correlation. The mind demands a tool for defining the *extent* of correlation shown by a vast number of pairs of measures on the two traits in question. A coefficient designed to measure relationship is valuable to the extent that it does this.

Our next problem therefore should be to show the common-sense significance of the correlation coefficient, and of the regression coefficients, and to indicate the relative degree to which they aid us in interpretation of our data. Suppose from the example given in Diagram 46, —

$$r = .48;\ \sigma_y = 1.26;\ \sigma_x = 0.89,$$

then,

$$r\frac{\sigma_y}{\sigma_x} = .68, \quad \text{and} \quad r\frac{\sigma_x}{\sigma_y} = 0.34.$$

Then

$$y = .68x \quad \text{and} \quad x = .34y.$$

In this problem we have three statements to aid us in the interpretation of the question: To what extent is ability in shop practice accompanied by ability in drawing, or *vice versa*. In the first place, we may use the value of the correlation coefficient $r = .48$. The question arises: What does this mean? Is there a direct relationship between the two abilities? If so, is there an indication of considerable relationship, little relationship, or no relationship? The sign of the correlation coefficient, which in this case is positive, answers the first question definitely. This positive sign means that any increase in one trait is accompanied by an increase in the other, and *vice versa*. Had the sign of r been negative, then an increase in, say x, would have been accompanied by a decrease in y, and *vice versa.*

To make clear the meaning of various *values* of r, suppose each series of measures had been ranked in order of size, as in Table 33. If the position of each measure were the same in both series (*i.e.*, if pupil A were first in both series, pupil B second in both series, pupil C third, etc., throughout), then the correlation between the two traits would be perfect and positive, and r would be $+ 1$. On the other hand, if the order of the pupils in the two series were exactly reversed (*i.e.*, the first pupil in one series should be the last in the other series, the second in one series should be the second last in the other series, etc.), then the correspondence ("correlation") again would be perfect but this time negative, and r would equal $- 1$. Again, if there should be no correspondence in the position of the measures in the two series, the value of r would be 0. Thus the value of r may range from -1 to $+1$. When between 0 and 1 it will "express a tendency, greater or less according to r's size, for measures above the mean position in one series to be above the mean position in the other series. When r is between 0 and -1, it will express a tendency, greater or less, according as r is numerically greater or less, for the measures *above* the mean position of one series to be *below* the mean position in the other, and conversely." The exact degree of relationship is commonly inferred from the relative size of the coefficient, r. Thus correlation may be spoken of as "high," "low," etc. It can be seen that the *definite* interpretation of correlation depends on the arbitrary placing of the limits of the values of r, which are to be called "high," "low," etc.

"High" and "low" correlation. This definition of limits depends largely on the personal experience of the person making the interpretation. For example, it has been common for certain educational investigators to arbitrarily interpret a coefficient of .25 as an indication of "high" positive correlation, and one of .40 as "very high." Others would interpret .25 as very low, and .50 as "marked" or "somewhat high." Certainly, our educational conclusions must be colored by our arbitrary definition of such a coefficient. The experience of the present writer in examining many correlation tables has led him to regard correlation as "negligible" or "indifferent" when r is less than .15 to .20; as being "present but low" when r ranges from .15 or .20 to .35 or .40; as being "markedly present" or "marked," when r ranges from .35 or .40 to .50 or .60; as being "high" when it is above .60 or .70. With the present limitations on educational testing few correlations in testing will run above .70, and it is safe to regard this as a very high coefficient.

The interpretation of the coefficient $r = .48$, in the above problem, would result in a general statement to this effect:

"There is *marked evidence* that abilities in shop practice and drawing accompany each other. Students *above the average* in one group will TEND to be *above the average* in the other. It is not known more specifically in what way the two abilities are centrally connected, or to what extent the presence of either one is an indication of the presence of the other." Except in the case in which the variability is the same, r does not enable us to foretell, for example, knowing the value of one trait, what, on the average, the value of the other will be. It does not enable us to say that for a given unit-change in abilities in shop practice, what changes should be expected, on the average, in drawing abilities.

A more complete method of describing relationship. This very vagueness in the possibility of definition of r leads us to turn to the more complete method of describing the relationship: namely, the equation of the line. Taking that, we now find that, for the regression of y on x, —

$$y = .68x,$$

and that for every unit deviation from the type of x (abilities in shop practice), *it is most probable* that there will be an accompanying deviation of .68 as much in y (abilities in drawing). The tendency, in the past, has been to stop the analysis of the data at this point, the conclusion being drawn that the two abilities are very closely related. It must be remembered, however, that there are *two regression lines*, one for the means of the columns and the other for the means of the rows. The former shows the deviation in y corresponding on the average to a unit deviation in the type of x, and the latter the deviation in x corresponding, on the average, to a unit deviation in the type of y. Thus, in our problem, $x = .34y$; *i.e.*, it is probable that a unit deviation in y will be accompanied by a deviation of .34 as much in x.

This explanation has made use of the "*deviation*" *formula*

(2). Using the formula (10) in which y and x are actual values instead of deviations, we can make this still clearer to the student. The equation now becomes

$$y - 85.55 = 48.1 \frac{1.26}{.89} (x - 85.25)$$

or, —

$$y - 85.55 = .68 (x - 85.25)$$

In this case, it must be remembered that x and y are actual values of the two traits, abilities in shop practice and drawing, and for y and x have been substituted the values of their respective means, $y = 85.55$; $x = 85.25$. Expressing the equation of the line of the means now enables us to assign values to one of the traits, say x, and compute the accompanying value of y. In Table 35, values decreasing by 5 have been assigned to x, and the y's computed. It will be noted that as each x decreases by 5 (90, 85, 80, 75, etc.), the corresponding decrease in the unit of y is $.68 \times 5 = 3.40$.

TABLE 35. REGRESSION OF x ON y

x	y
95	92.18
90	88.78
85	85.38
80	81.98
75	78.58
70	75.18
60	68.38

TABLE 36. REGRESSION OF y ON x

y	x
95	88.46
90	86.76
85	85.06
80	83.36
75	81.66
70	79.96
60	76.56

This should make clear the statement made above that a given deviation in x would be accompanied by .68 as much change in y. In the same way Table 36 gives corresponding values for y and x computed from the regression equation of y on x: —

$$x - 85.25 = .34\ (y - 85.55).$$

The effect of the smaller regression coefficient (.34 instead of .68), is now seen in the relative values of y and x. As y decreases steadily by 5 units, x decreases by only 1.70 units $(.34 \times 5 = 1.70)$. Reference to Diagram 46 will reveal the way in which differences in relationship between the two traits are partially described by the "slope" of the line of the means. The plotting of the equation of the line of the means of the *columns*, —

$$y = .68x$$

gives a line of considerable steepness, CC′. For given changes in x we have nearly proportional changes in y. The plotting of the equation of the line of the means of *rows*, —

$$x = .34y,$$

gives a line much flatter in slope, R_1R_2. For given changes in y we have much smaller changes in x.

(*c*) **How to plot the line of the means.** We are now in a position to *draw the line of regression* on our correlation table. There are two methods by which this may be done. The first is the rough method of drawing, from inspection of the mean points of the columns and rows, a line which most closely approximates them. This can be done by laying a celluloid triangle, or a thread over the table, and adjusting it by eye until it most closely fits the mean points of the columns and rows. The line may be drawn accurately, however, by first computing the equations of the lines of the means. Values may then be assigned to x, and corresponding values of y can then be computed, exactly as in Tables 35 and 36. Since a straight line can be plotted from any two of its points, we can draw the line by plotting any two of the pairs of coördinates, x and y. For example, in Table 39, the line CC' is determined by connecting C' (which was plotted from $x = 90$, $y = 88.88$) and C (plotted from $x = 60$, $y = 68.38$). The remaining points of the table will fall on the same line, since their coördinates have been computed from the equation of this line.[1]

3. Computation of the correlation coefficient and the regression coefficients

It is now clear that statistical methods can supply us with a tool for estimating *relationship* in terms of the most probable values of two concurrently changing quantities. The determination of the law of relationship must lead to *the computation of the regression coefficients.* This in turn demands the computation of the correlation coefficient r, which, in itself, will throw *some* light on the status of relationship. There are two principal steps in the computation of these coefficients: (1) the tabulation of the correlation table; (2) the computation of three devices, σ_x, σ_y, and r, with the consequent substitution of these values in the regression equations.

(*a*) The first step: the tabulation of the correlation table. The foregoing pages have made it clear that complete interpretation of correlation demands the tabulation of each of the pairs of measures in the correlation table. The steps in the tabulation may be conveniently listed as follows: —

(1) Decide on the *size* and *position* of the class-intervals in each distribution. This should be done in accordance with the principles laid down in Chapter IV, in the discussion

---

[1] The more refined methods of fitting lines to plotted data, involving, as they do, the theory of curve-fitting, will not be taken up in this work. In the bibliography at the end of the book complete directions are given the mathematically trained student for finding the literature.

of the classification of data in a single frequency distribution. The student must understand that he is *now to tabulate pairs of measures* which occur in two sets of class-intervals at the same time.

(2) Write the limits of these class-intervals along the two axes of the table, assigning one trait to y and the other to x. Lay off these limits from a zero point, supposed to be at the bottom and left of the table, as in Diagram 45; *e.g.*, 61–65, 66–70, etc., from bottom up, and 71–75, 76–80, etc., from left to right.

(3) Having the original measures arranged in parallel series, as in Table 33, tabulate these pairs of measures in the appropriate rectangle in which they fall. It will be helpful to have the *y-series* on the left, and the *x-series* on the right in this pairing of the measures. The caution stated in Chapter IV to define carefully the limits of class-intervals should be kept in mind in this work. More errors are made in the original tabulation of

		Ability in Shop Practice				
		71_ 75	76_ 80	81_ 85	86_ 90	91_ 95
Ability in Drawing	100_ 96			/	//	/
	95_ 91		//	𝍸 ///	𝍸 𝍸 𝍸 𝍸 //	𝍸
	90_ 86		///	𝍸 𝍸 𝍸 𝍸 /	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 /	𝍸 ///
	85_ 81	//	𝍸 ////	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸	𝍸 𝍸 𝍸 /	/
	80_ 76	//	𝍸	𝍸 𝍸 𝍸	𝍸 ///	
	75_ 71	/	𝍸 /	////		
	70_ 66		/	/	/	
	65_ 61	/				

DIAGRAM 45. TO ILLUSTRATE THE FIRST STEP IN PLOTTING A CORRELATION TABLE

Checking pairs of measurements in appropriate compartments of the table.

the correlation table than in any other one aspect of the work. The tabulation is illustrated in Diagram 45.

(4) The pairs of measures having been checked on the table in pencil, next replace the checking by numbers, to give a table similar to Table 37.

TABLE 37. TO ILLUSTRATE ANOTHER PHASE OF THE SECOND STEP IN THE TABULATION OF A CORRELATION TABLE

Ability in Shop Practice

		71–75	76–80	81–85	86–90	91–95
Ability in drawing...	100–96			1	2	1
	95–91		2	8	22	5
	90–86		3	21	31	8
	85–81	2	9	30	16	1
	80–76	2	5	15	8	
	75–71	1	6	4		
	70–66		1	1	1	
	65–61	1				

(*b*) The second step: the computation of the coefficient of correlation r and the regression coefficients, b_1 and b_2. Our task is to compute r from the formula

$$r = \frac{\Sigma xy}{N\sigma_x\sigma_y}$$

and b_1 and b_2 from the formulæ —

$$b_1 = r\frac{\sigma_y}{\sigma_x}$$

$$b_2 = r\frac{\sigma_x}{\sigma_y},$$

the final equations of the lines of regression being —

$$y = r\frac{\sigma_y}{\sigma_x}x \text{ (for the regression line of the columns),}$$

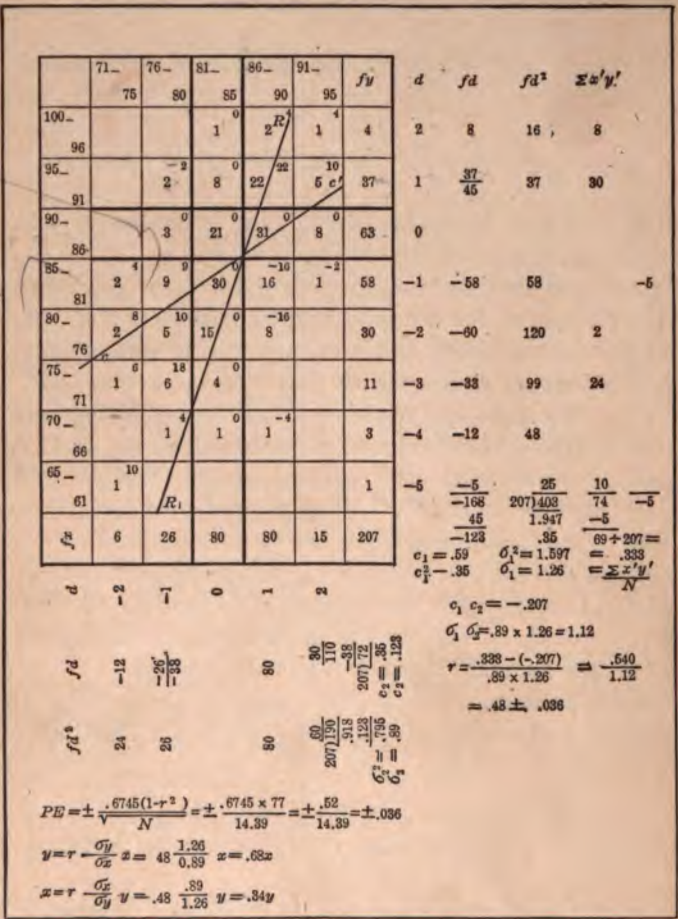
and

$$x = r\frac{\sigma_x}{\sigma_y}y \text{ (for the regression line of the rows).}$$

The work may be made clear by first listing the steps in the computation of r. The formula requires us to find the two standard deviations, σ_y, for the total frequency columns of the y's, and σ_x for the total frequency rows of x's. The student's first difficulty in understanding the computation will be in comprehending clearly that *σ_y and σ_x are the standard deviations of the total frequency distribution of the columns and rows*. Thus, in Diagram 46, the column and row headed f_y and f_x mean respectively "total frequency of the y's" and "total frequency of the x's." Thus, the standard deviations, σ_y and σ_x are found from these two frequency distributions exactly as described in Chapter VI. Furthermore, the short method of computation can be applied to the two distributions to cut down greatly the labor of computation, not only for the standard deviations but also for Σxy.

Steps in the computations. The entire steps in the computation are as follows (compare Diagram 46 for illustrative references): —

1. Total the measures in each distribution, giving N.
2. Estimate the class-interval which contains the mean, *e.g.*, 86–90 for the y's; 81–85 for the x's.
3. Tabulate the deviation in unit intervals, of the mid-value of each class-interval from that of the estimated mean, 1, 2, 3, etc., -1, -2, -3, etc.
4. Multiply each frequency by its respective deviation; *e.g.*, for the y's, $4 \times 2 = 8$, $37 \times 1 = 37$, etc., for the x's, $6 \times -2 = -12$, $26 \times -1 = -26$, etc.



	71– 75	76– 80	81– 85	86– 90	91– 95	*fy*	*d*	*fd*	*fd*²	Σ*x*′*y*′	
100– 96			1^{0}	2^{4} *R*	1^{4}	4	2	8	16	8	
95– 91		2^{-2}	8^{0}	22^{22}	5^{10} *c*	37	1	$\frac{37}{45}$	37	30	
90– 86		3^{0}	21^{0}	31^{0}	8^{0}	63	0				
85– 81	2^{4}	9^{9}	30^{0}	16^{-16}	1^{-2}	58	−1	−58	58		−5
80– 76	2^{8} *c*	5^{10}	15^{0}	8^{-16}		30	−2	−60	120	2	
75– 71	1^{6}	6^{18}	4^{0}			11	−3	−33	99	24	
70– 66		1^{4}	1^{0}	1^{-4}		3	−4	−12	48		
65– 61	1^{10}	*R*₁				1	−5	−5	25	10	
fx	6	26	80	80	15	207		−168	207)403	74	−5
								45	1.947	−5	
								−123	.35	69 ÷ 207 =	
d	−2	−1	0	1	2			$c_1 = .59$	$\sigma_1^2 = 1.597$	= .333	
fd	−12	−26 −38		80	30 110 −38 207)72 $c_2 = .35$ $c_2^2 = .123$			$c_1^2 - .35$	$\sigma_1 = 1.26$	$= \frac{\Sigma x'y'}{N}$	
*fd*²	24	26		80	60 207)190 .918 .123 $\sigma_2^2 = .795$ $\sigma_2 = .89$						

$c_1\ c_2 = -.207$

$\sigma_1\ \sigma_2 = .89 \times 1.26 = 1.12$

$$r = \frac{.333 - (-.207)}{.89 \times 1.26} = \frac{.540}{1.12}$$

$$= .48 \pm .036$$

$$PE = \pm \frac{.6745(1-r^2)}{\sqrt{N}} = \pm \frac{.6745 \times 77}{14.39} = \pm \frac{.52}{14.39} = \pm .036$$

$$y = r \frac{\sigma_y}{\sigma_x} x = 48 \frac{1.26}{0.89} x = .68x$$

$$x = r \frac{\sigma_x}{\sigma_y} y = .48 \frac{.89}{1.26} y = .34y$$

DIAGRAM 46. TO ILLUSTRATE COMPUTATION OF THE CORRELATION COEFFICIENT AND THE REGRESSION COEFFICIENTS FOR THE CASE OF LINEAR REGRESSION

(Adapted from form used by Dr. H. L. Rietz, of the University of Illinois.)

5. Find the algebraic sum of such fd's, *e.g.*, $\Sigma fd_y = -168 + 45 = -123$; $\Sigma fd_x = 110 - 38 = 72$.
6. Divide Σfd by the number of cases, N, to give the correction c; *e.g.* —

$$c_y = \frac{-123}{207} = -.59; c_x = \frac{72}{207} = .35.$$

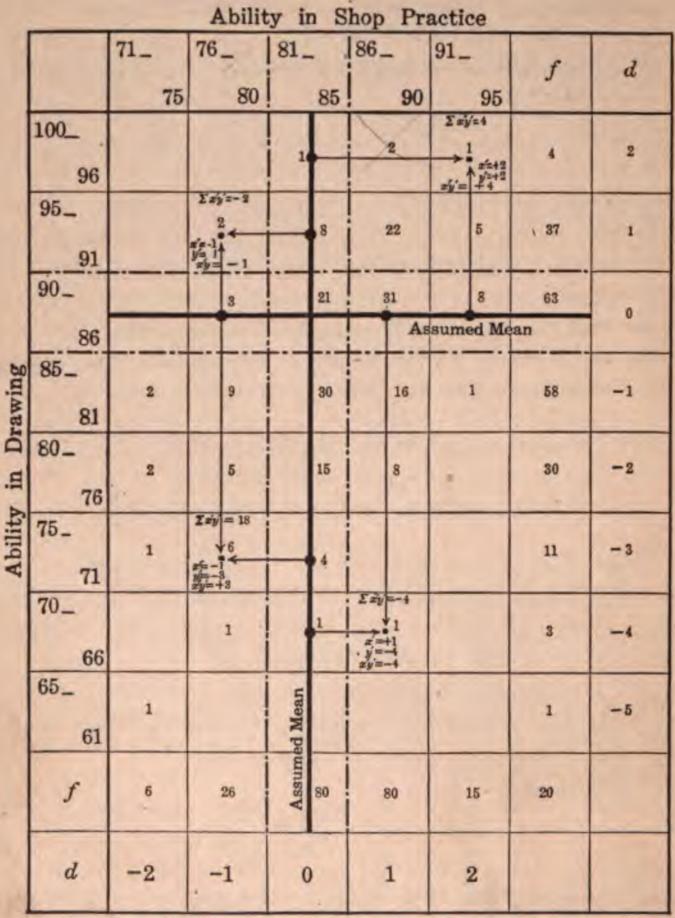
7. Square the corrections; *e.g.*, $c_y^2 = .35$; $c_x^2 = .123$.
8. Multiply each fd by d, its corresponding deviation, to give column headed fd^2; *e.g.*, $fd^2_y = 16, 37, 0, 58, 120$, etc., $fd^2_x = 24, 26, 0, 80, 60$.
9. Find the sum of the fd^2; *e.g.*, $\Sigma fd^2_y = 403$; $\Sigma fd^2_x = 190$.
10. Divide this sum by N, to give S^2 the square of the standard deviation of each distribution *around the assumed mean; e.g.*, $S_y^2 = 1.947$; $S_x^2 = .918$.
11. Subtract the square of the correction from S^2; *e.g.*, $\sigma_y^2 = 1.947 - .35 = 1.597$; $\sigma_x^2 = .918 - .123 = .795$.
12. Find the square root of σ^2 giving σ; *e.g.*, $\sigma_y = 1.26$; $\sigma_x = .89$.

Note that these standard deviations are expressed in *units of class-intervals of 1*, and that to find the correlation coefficient, r, *they may be left in these units, provided* $\Sigma x'y'$ *is computed in the same units.* It will cut down the labor of computation greatly to do this. Note, furthermore, that the above twelve steps merely restate the steps in the computation of σ as given in Chapter VI.

The formula

$$r = \frac{\Sigma x'y'}{N\sigma_x\sigma_y}$$

next demands that we compute the product-sum of the corresponding pairs of deviations from their respective means x_1y_1, x_2y_2, x_3y_3 for every point in the correlation table. Diagram 47 will make clear what is wanted. The two measures in the compartment $y=96-100$, $x=86-90$, each deviate from the mean of the x's, *i.e.*, from $\bar{x}$ by 1 class-



	71–75	76–80	81–85	86–90	91–95	*f*	*d*
100–96			1	2	1	4	2
95–91		2	8	22	5	37	1
90–86		3	21	31	8	63	0
85–81	2	9	30	16	1	58	−1
80–76	2	5	15	8		30	−2
75–71	1	6	4			11	−3
70–66		1	1	1		3	−4
65–61	1					1	−5
f	6	26	80	80	15	20	
d	−2	−1	0	1	2		

DIAGRAM 47. A PRODUCT-MOMENT DIAGRAM

To illustrate the computation of $\Sigma x'y'$. For the data of Diagram 46. The computation is illustrated graphically for one compartment in each quadrant. x' and y' are deviations (or "moments") of the mean of the compartment from the respective assumed means of the table.

interval $(=x')$ and from $\bar{y}$ by 2 class-intervals $(=y')$. That is, for each of these two measures $x'y' = 1 \times 2$. For the measures in the compartment $y = 96\text{–}100$, $x = 91\text{–}95$, $x'y' = 2 \times 2$; for the compartment $y = 76\text{–}80$, $x = 96\text{–}90$, $\Sigma x'y'$
$\Sigma \quad x' \qquad y'$
$= 8\ [1 \times -2] = -16$. Note carefully that *the signs of the deviations must be taken account of*. These signs are now determined by noting whether the measure in question is greater than or less than the mean of the total distribution. A measure *greater than* the mean will deviate *positively;* one less than the mean will deviate *negatively*. To expedite the work of the student the correlation table should be divided into *four quadrants*, as follows: —

$x = -$ $y = +$	$x = +$ $y = +$
$x = -$ $y = -$	$x = +$ $y = -$

If the class-intervals have been laid off as suggested from left to right, and from bottom upward, the quadrants, with the signs of x and y, are as just given.

Now, to compute $\Sigma x'y'$ for the whole table, going from compartment to compartment and summing the product of the pairs of measures, as shown above, will be a very laborious task. The labor may be shortened very much by summing the x *deviations of all the measures in one row, and multiplying* $\Sigma x'$ *once for all by* y'. This method recognizes that *all the measures in a given row, e.g.*, 2, 8, 22, 5, in row 91–95, *have the same* y', namely, $+1$. Treating the material in this way enables us to compute the deviations mentally and very rapidly. With this explanation we are now ready for step 13 in the computation of r.

13. Compute $\Sigma x'y'$, by finding the sum of the deviations of the measures in a particular row *from the mean of the x's of the whole table* ($\bar{x}$). This gives $\Sigma x'$. Multiply $\Sigma x'$ by y', the deviation of this particular row from $\bar{y}$ the mean of the y's *of the whole table*. This gives $\Sigma x'y'$, *which is the product-sum of the deviations about the two assumed means.*

TABLE 38. COLUMNS CORRESPONDING TO ROW 96–100

Row 96–100	81–85	86–90	91–95	Total $\Sigma x'$
$x' =$	0	+1	+2	
$n =$	1	2	1	
$\Sigma x' =$	0	+2	+2	+4
$y' =$				+2
$\Sigma x'y' =$				+8

TABLE 39. COLUMNS CORRESPONDING TO ROW 76–80

	71–75	76–80	81–85	86–90	*Totals* $\Sigma x'$
$x' =$	−2	−1	0	+1	
$n =$	2	5	15	8	
$\Sigma x' =$	−4	−5	0	+8	−1
$y' =$					−2
$\Sigma x' y' =$					+2

The computation, presented here in tabular form, can be done mentally. In setting down the results of the $\Sigma x'y'$ for each row, as 8, 30, etc., in Table 39, it may be more accurate for the beginning student to tabulate both the positive and negative $\Sigma x'$ separately, summing them both separately to give the algebraic sum of the deviations. In the accompanying problem, Diagram 46, the work has all been done mentally, the algebraic sum of the $\Sigma x'y'$ being tabulated in one column. This gives $\Sigma x'y' = .69$.

This product-sum is for deviations computed from the two *assumed* means, not the true means. Therefore, just as the means are in error by a correction c_y, or c_x, so each deviation on y and on x is in error by the same amount. Thus, since we must apply corrections to find the true means and the true standard deviations, so we must apply a correction to find Σxy, the product-sum of the deviations about the true means. This means that we must multiply c_x and c_y together to get this correction. It has been shown that the formula for r, by this short method of computing the terms about the assumed mean, is: —

$$r = \frac{\dfrac{\Sigma x'y'}{N} - c_x c_y{}^*}{\sigma_x \sigma_y}$$

* Let E_x and E_y represent the estimated means of the two series, and c_x and c_y be corrections to be applied to the estimated means to get the true means. Then the True Means, M_x and M_y are respectively $M_x = E_x + c_x$ and $M_y = E_y + c_y$.

Let x and y be deviations from the True Means, M_x and M_y.

Let x' and y' be deviations from the Estimated Means, E_x and E_y.

Thus, $x' = x + c_x$ and $y' = y + c_y$.

$$\text{Therefore, } \Sigma x'y' = \Sigma(x + c_x)(y + c_y)$$
$$= \Sigma xy + c_y \Sigma x + c_x \Sigma y + \Sigma c_x c_y.$$

Now, since Σx and Σy (the sum of the x and y deviations *from the TRUE MEAN*) each $= 0$, then

$$\Sigma x'y' = \Sigma xy + \Sigma c_x c_y, \text{ or } \Sigma xy = \Sigma x'y' - \Sigma c_x c_y$$

or, substituting this expression in the equation

$$r = \frac{\Sigma xy}{N\sigma_x \sigma_y}$$

we get
$$r = \frac{\Sigma x'y' - Nc_x c_y}{N\sigma_x \sigma_y} = \frac{\dfrac{Sx'y'}{N} - c_x c_y}{\sigma_x \sigma_y}$$

(Adapted from H. L. Rietz; Bulletin no. 148, University of Illinois Agricultural Experimentation Station. 1910.)

14. Divide $\Sigma x'y'$ by N. In the problem in Diagram 46,

$$\frac{\Sigma x'y'}{N} = .333.$$

15. Multiply c_x by c_y; *e.g.*, $c_x c_y = -.207$.

16. Subtract $c_x c_y$ from $\frac{\Sigma x'y'}{N}$; *e.g.*, .540.

17. Divide

$$\frac{\frac{\Sigma x'y'}{N} - c_x c_y}{\sigma_x \sigma_y}$$

giving r, *the coefficent of correlation.*

$$r = .48.$$

18. The regression coefficients can now be computed. The regression of y on x is —

$$b_1 = r\frac{\sigma_y}{\sigma_x} = .48\,\frac{1.26}{.89} = .68.$$

$$b_2 = r\frac{\sigma_x}{\sigma_y} = .48\,\frac{.89}{1.26} = .34.$$

That is, divide the standard deviations, one by the other, and multiply by r.

19. Write the equations of the two lines of relationship,

$$y = r\frac{\sigma_y}{\sigma_x}x \text{ and } x = r\frac{\sigma_x}{\sigma_y}y.$$

That is, $y = .68x$, and $x = .34y$.

20. These lines may now be plotted accurately on the table by assigning values of x and computing corresponding values of y and *vice versa*.

(*c*) **Reliability of the correlation coefficient.** The mere statement of the value of a correlation coefficient, taken alone, is not sufficient evidence of relationship between the two traits. Having computed r (say, $r = .35$) we must determine the *reliability of our coefficient*. This question arises: If we should continue to take "samples" from the general population, under the same conditions with which we took our original sample, would such successive "*samples*" continue to give the same correlation coefficient? More concretely: Suppose we wish to find the relationship between ability to spell and ability to add in a very large school population, say, 20,000 pupils. Suppose that we have tested the two abilities, adequately, in a "random" sample of 200 pupils from this population. The correlation coefficient r proves to be $+ .35$, leaving with us a belief that the two abilities accompany each other rather generally. We now ask: If we continue to take, at random, samples of *200* pupils each from the entire 20,000 children, will r continue to be approximately .35? Or, could r fluctuate considerably merely from conditions of sampling?

There are two methods of solving this problem, — the first the practical, but laborious method of continuing to take successive samples, making the group larger and larger until the coefficient does become stable. This common-sense method requires too much labor in the collection and treatment of data to be practically useful.

The second or statistical method, and the one universally used, is to turn to the question of "chance" and determine the probability that such a coefficient will remain stable. It is clear, therefore, that the determination of *reliability* of a correlation coefficient, like that of a mean, or of a standard deviation, must depend on the "*normality*" of the distributions in question. There are two distinct questions involved: (1) Do the original data, when plotted, approximate a normal probability distribution? (2) If so, what is the ratio between the size of the coefficient and the size of the probable error, *P.E.*?

Thus, the first step should be to plot the data, or at least to note whether they show fair concentration near the middle of the scale. It must be remembered that, with a small number of cases (*i.e.*, certainly when N is less than 30), the possibility of resemblance of the distribution to normality will be very doubtful. Under such cases, we do not know how much the probable error will be. If the distribution can be said to *resemble* normality, then recourse may be had to the *P.E.* to enable us to estimate the *probable stability* of the coefficient.

It was pointed out in Chapter VIII that the *P.E.* of r could be found from the formula —

$$P.E._{\cdot r} = .67449 \frac{1 - r^2}{\sqrt{N}}$$

Interpreted in words, this means that the chances are even (1 to 1) that the true value of r lies within the limits —

$$r \pm P.E., \text{ or } r \pm .67449 \frac{1 - r^2}{\sqrt{N}}\text{*}$$

Thus, from the relative sizes of r and *P.E.* we can state limits outside of which it is very improbable that the true value will fall. For example, statistical practice has tended to set the criterion that the correlation coefficient should be at least 3 times as large as the probable error; this, largely on the ground that it is very improbable that the true value of r falls outside $r \pm 3$ *P.E.* More conservative practice insists upon r being 4 times *P.E.*

Thus, when N is not very small, the computation of r should always be supplemented by the computation of *P.E.*, and r should be reported in the form: $r \pm$ *P.E.* For example, in the problem of Diagram 46, .48 ± .04.

---

* At this point the student is referred again to the discussion of the probability curve in Chapters VII and VIII.

Table for determining *P.E.* It should be noted that *P.E.* is directly a measure of *un*reliability. The formula shows that the unreliability increases as N, the number of cases, grows smaller. Conversely the coefficient r grows more reliable as N increases, but in proportion to the square of the number of cases. Thus, to *double* the reliability of a coefficient, we must take 4 times the number of cases. To triple the reliability of r, *i.e.*, to reduce the *P.E.* to one third of its present value, we must take 9 times the number of cases. This is illustrated concretely in Table X, in the Appendix, a table which gives at once the values of the *P.E.* for various values of r and n. Thus, if $r = .3$ and $N = 25$, the $P.E. = .1228$, nearly one half of r, which means doubtful reliability. The table tells us that in order to double the reliability, making *P.E.* .0614, we must take 100 cases (4 × 25).

In this connection an important practical question faces every investigator in the collection of educational data. How many measures must be collected in order to insure a coefficient which is statistically reliable? This amounts to asking: How can we select a random sample? The criterion of the probable error enables us to answer such a question in a rough way as follows: Assuming the worst possible condition as to correlation, *i.e.*, assuming r to be small, .1, .2, or .3 (unless as in rare cases, the investigator can estimate the coefficient and knows it to be high), determine from Table X the number of cases that are necessary to give *P.E.* not more than one third to one fourth of r. For example, if r is estimated in advance to be as low as .2, the investigator ought to take at least 100 cases to insure a sufficiently reliable r. The taking of a sample on such grounds satisfies ONLY this statistical criterion of probability. It should be noted, furthermore, that the value of N which is assigned should refer to the *smallest* group for which correlations are to be computed. If r were .4 or more, then 25 cases would give sufficient reliability to the coefficient, according to present practice in the interpretation of correlation coefficients and probable errors. With such a small number of cases, however, it is clear that the criterion of the probable error cannot be used. When N is so small "that certain higher powers of its reciprocal cannot be neglected in comparison with the rest of the expression involving them, the values (of the probable error) cannot be used. For such cases no theoretical formulæ have hitherto been devised." [1]

4. *Computation of straight line relationship without the tabulation of the correlation table*

Short method. It is possible to turn the product-moment formula —

$$r = \frac{\Sigma xy}{N\sigma_x\sigma_y}$$

into the expression

$$r = \frac{\Sigma x \cdot y}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$$

---

[1] Brown, W., *Essentials of Mental Measurement*, p. 61.

W. Brown further cites an *empirical* investigation on the determination of the reliability of r for small numbers of cases, 4, 8, and 30 respectively, "taken from a total population of 3000 pairs of measurements (height, and left middle-finger measurements of 3000 criminals: 'real' correlation, .66) ... Correlation results, for real value of

	$r = .66$, were
Samples of 4	$.561 \pm .011$
" " 8	$.614 \pm .065$
" " 30	$.6609 \pm .0067$

Hence it may be concluded that, although in the case of such small samples as 4 or 8 the ordinary formula for the *P.E.* of r gives much too low a value, yet in the case of as many as 30, the formula applies with tolerable accuracy. We must, however, bear in mind that this result has only been proved (empirically) to hold in the single case where the actual correlation was .66."

TABLE 40. TO ILLUSTRATE COMPUTATION OF r WITHOUT TABULATION OF THE CORRELATION-TABLE

Individual	*Score in I*	*Score in II*	x *diff. of scores in I from average*	y *diff. of scores in II from average*	x^2	y^2	xy
1......	15	10	−4	−3	16	9	+12
2......	15.5	10	−3.5	−3	12.25	9	+10.5
3......	16	6	−3	−7	9	49	+21
4......	17.5	10	−1.5	−3	2.25	9	+ 4.5
5......	17.5	11	−1.5	−2	2.25	4	+ 3.0
6......	17.5	18.5	−1.5	+5.5	2.25	30.25	− 8.25
7......	18.5	11	− .5	−2	.25	4	+ 1
8......	19.5	13	+ .5	0	.25	0	0
9......	20.5	10	+1.5	−3	2.25	9	− 4.5
10......	20.5	13	+1.5	0	2.25	0	0
11......	20.5	20	+1.5	+7	2.25	49	+10.5
12......	22	17.5	+3	+4.5	9	20.25	+13.5
13......	23.5	16	+4.5	+3	20.25	9	+13.5
14......	24	18	+5	+5	25	25	+25
Average.	19	13			105.5	226.5	101.75

$$r = \frac{\Sigma x \cdot y}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}} = \frac{101.75}{\sqrt{105.5 \times 226.5}} = \frac{101.75}{154.6} = 65.8$$

$$\left(\text{or } \frac{\text{sum of the products of } x \text{ and } y}{\text{square root of (the sum of } x^2 \times \text{ the sum of } y^2)}\right)$$

It has been common practice among educational workers to compute r by the use of this formula, without the necessary tabulation of the correlation table and the determination of the linearity of regression. It has been shown in the foregoing pages that, in order to be able to apply the product-moment formula the data in question must reveal straight line relationship, — because *r is a term in the equation of the straight line which "best fits" the means of the table.* However, when N is very small, it is questionable whether *any* method of correlation gives very reliable results. For that reason it may be desirable to have available approximate or short methods of computing correlation, for purposes of rough preliminary examination of the data. We shall take up in later sections the "*rank*" and *fourfold methods* of doing this. At this point we should refer to the use of the short method for finding correlation, applicable when regression is linear.

Table 40 illustrates[1] the method in detail.

II. THE CASE OF NON-LINEAR RELATIONSHIP

When the line of the means is not a straight line. It is clear that if the means of the correlation table do not accord fairly well with a straight line, the product-moment formula for r and the regression equations of the "best fitting lines" cannot be used to describe the relationship between the two traits under consideration. At the same time, we must note that there may be a decided relationship between two traits, even though the line of the means is not a straight line. For example, Diagram 48 presents a case of high correlation, the use of r for which leads to distinctly incorrect conclusions. In this diagram[2] the product-moment formula was used, giving $r = - .47$. The correlation is actually $- .83$, when computed by proper methods ($\eta = - .83$). Mere inspection of the table leads to the conclusion that it is *not* permissible to describe such a table by the equation of a straight line.

It is evident, therefore, that we need a method of computing a coefficient for those kinds of relationship in which the means of the table do not fall approximately on a straight line. Note then, that we shall seek a method of describing

---

[1] Quoted from Freeman, F. N., *Experimental Education*, p. 178.

[2] Monroe, W. S., *The Cost of Instruction in Kansas High Schools.* Studies by the Bureau of Educational Measurements and Standards, No. 2. (Kansas State Normal School, Emporia, Kansas, 1915.)

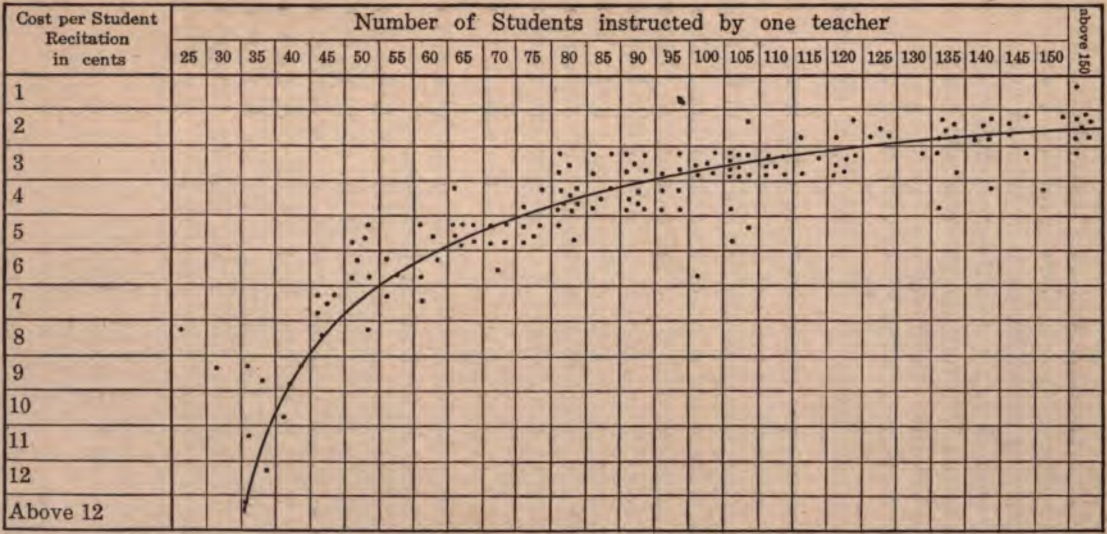


DIAGRAM 48. RELATION BETWEEN COST-PER-STUDENT-RECITATION IN ENGLISH AND THE NUMBER OF PUPILS INSTRUCTED BY ONE TEACHER IN 148 KANSAS HIGH SCHOOLS

(Date from W. S. Monroe, 1915.) To illustrate use of correlation-ratio (η) and linear regression methods.

relationship which, as before, will treat the separate columns of the table in terms of their arithmetic means. We shall be interested in finding how a deviation from the means of one trait (measured on, say, the x-axis) corresponds *on the average*, to a *unit deviation* from the mean of the other trait, (measured on the other, the y-axis).

The product-moment method involves finding the product of the separate ratios of $\frac{x}{\sigma_x}$, $\frac{y}{\sigma_y}$, that is, of the deviations of every measure of the table from the mean of its distribution, *measured in units of the standard deviation of the x's or y's of the whole table.* We find the amount that each measure differs from the mean of its distribution. This is x for the x-measures, and y for the y-measures. These x and y deviations can be made comparable by dividing each one by its respective standard deviation as a unit. Thus, in computing the correlation coefficient, "r," we *measure each deviation x or y* in units of its corresponding standard deviation σ_x or σ_y, *i.e.*, $\frac{x}{\sigma_x}$ and $\frac{y}{\sigma_y}$. r is the ratio between the sum of the x deviations times the y deviations, each measured in terms of its standard deviation.

Professor Pearson has suggested that the non-linear tables may be treated by finding the ratio of the standard deviation of the arithmetic means of each of the columns (or rows) of the table to the standard deviation of the whole table itself (the σ_y for the columns and σ_x for the rows of the table). In symbols this means (letting the general expression be called the "correlation-ratio" $=\eta$)

$$\eta = \frac{\Sigma}{\sigma_y} = \frac{\sqrt{\frac{S(n_x(\bar{y}_x - \bar{y})^2)}{N}}}{\sigma_y}$$

where,

n_x = total number of measures in any column;
$\bar{y}_x$ = the arithmetic mean of any column;
$\bar{y}$ = the arithmetic mean of all the y's in the table;
N = total number of measures;
σ_y = the standard deviation of all the y's in the table.

This is equivalent to saying

$$\Sigma = \sqrt{\frac{S\,[n_x(\bar{y}_x - \bar{y})^2]}{N}} =$$

the standard deviation of the means of the columns of the table. For, each $(\bar{y}_x - \bar{y})$ equals the difference between the arithmetic mean of a column $(\bar{y}_x)$ and the arithmetic mean of the total frequency obtained from all the columns in the table. That is, each $(\bar{y}_x - \bar{y})$ is the "deviation" of the mean of a column from the mean of all the y's in the table. These are each squared and weighted by their corresponding frequencies, n_x. Thus, it can be seen that the above formula is of the usual form of the standard deviation: —

$$\sigma = \sqrt{\frac{\Sigma f d^2}{N}}$$

That is, in the above symbolism, n_x is equivalent to f, the frequency; $\bar{y}_x - \bar{y}$; is equivalent to d; S is equivalent to Σ.

Diagram 49 is supplied to make clear the use of the symbolism: Table 41 illustrates the method in detail.

Summary of process. We may summarize the process of computing the correlation-ratio by listing the following steps. —

1. Tabulate correlation table in exactly the same way as in computing r.
2. Sum the columns (n_x) or the rows (n_y) of the table. (These correspond to f's in the computation of r.)
3. Compute the *arithmetic mean* of all the y's in the table. Call this $\bar{y}$. (This is done in exactly the same way as in com-

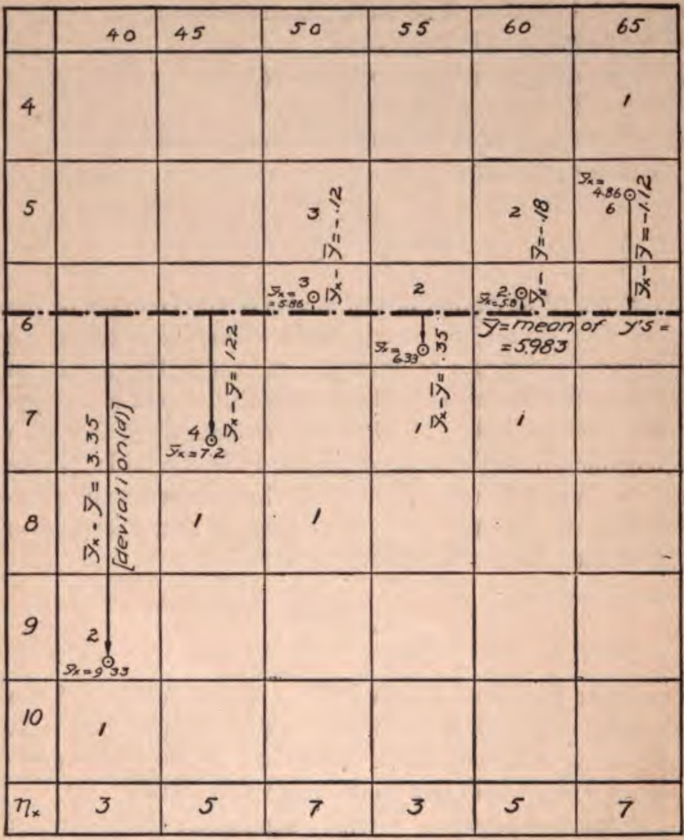


DIAGRAM 49. ABSTRACT FROM TABLE 41

To illustrate the fact, that, in the computation of the Correlation-Ratio (η), ($\bar{y}_x - \bar{y}$) is the deviation (d) of the mean of the column from the mean of the table.

puting the mean of any frequency distribution (see Chapter V). The distribution to use in this case is that of the total frequency column, headed n_y. The short method should be used, as before, using units of class-intervals instead of the original units.

TABLE 41. CORRELATION BETWEEN COST OF INSTRUCTION PER PUPIL RECITATION AND THE NUMBER OF PUPILS TAUGHT BY ONE TEACHER. TO ILLUSTRATE COMPUTATION OF CORRELATION RATIO FOR NON-LINEAR TABLES

		Number of Students Instructed by one Teacher															
		25	30	35	40	45	50	55	60	65	70	75	80	$N_y = f$	d	fd	fd²
Cost per Student Recitation in cents	3												3	3	−2	−6	12
	4									1		2	7	10	−1	−10	10
	5						3		2	6	4	4	2	21	0	0	0
	6						3	2	2		1			8	1	8	8
	7					4		1	1					6	2	12	24
	8	1				1	1							3	3	9	27
	9		1	2	2									5	4	20	80
	10				1									1	5	5	25
	11			1										1	6	6	36
	12			1										1	7	7	49
	13			1										1	8	8	64
$N_x = f$		1	1	5	3	5	7	3	5	7	5	6	12	60		59	335
Means $\bar{y}_x$		8	9	10.8	9.33	7.2	5.86	6.33	5.8	4.86	5.2	4.67	3.92				
$\bar{y}_x - \bar{y}$		2.02	3.02	4.82	3.35	1.22	−0.12	.35	−.18	−1.12	−0.78	−1.31	−2.06				
$(\bar{y}_x - \bar{y})^2$		4.08	9.12	23.23	11.22	1.49	.014	.12	.032	1.254	.608	1.716	4.244				
$N_x(\bar{y}_x - \bar{y})^2$		4.08	9.12	116.15	33.66	7.45	.01	.46	.16	8.78	3.04	10.3	50.93				

$C\bar{y} = \frac{59}{60} = .9833$

$C\bar{y}^2 = .967$

$\frac{\Sigma fd^2}{N} = \frac{335}{60} = 5.5833$

$\sigma y^2 = 5.58 - .97 = 4.61$

$\sigma y = \sqrt{4.61} = 2.15$

$\bar{y} = \text{Assumed Mean} + C\bar{y} = 5 + .983 = 5.983$

$S\,[N_x\,(\bar{y}_x - \bar{y})^2] = 244.13$

$$\Sigma = \sqrt{\frac{[N_x\,(\bar{y}_x - \bar{y})^2]}{N}} = \sqrt{\frac{244.13}{60}} = \sqrt{4.07} = 2.017$$

Correlation Ratio, $\eta = \frac{2.017}{2.15} = .94$

$$P.E = .6745\,\frac{1-\eta^2}{\sqrt{N}} = .6745\,\frac{1-.8836}{\sqrt{60}} = \frac{.0785}{7.746} = .01+$$

4. Compute the arithmetic mean of the y's *in each column of the table.* Call each of these $\bar{y}_x$. (Each of these can be left in the form of the *correction to the true mean,* or the difference between the true and assumed means, provided $\bar{y}$ is expressed in the same way. To do this will cut down the arithmetic labor somewhat.)
5. Compute the square of the standard deviation of the y's in the whole table; call this σ_y^2. (As in step 3, use the total frequency distribution, headed n_y.)
6. Subtract the arithmetic mean of the whole table $\bar{y}$, from the arithmetic mean of each column, $\bar{y}_x$, to find the amount of deviation of the mean of each column from the mean of the table. That is, perform the operation $(\bar{y}_x - \bar{y})$ for each column of the table. (This corresponds to finding d in the case of the computation of the standard deviation of any distribution.)
7. Square each of these deviations $(\bar{y}_x - \bar{y})$, giving $(\bar{y}_x - \bar{y})^2$.
8. Weight each of the deviations (squared) by the number of cases occurring in each column; that is, multiply each $(\bar{y}_x - \bar{y})^2$ by its corresponding n_x. (This corresponds to finding fd^2 in the common standard deviation formula.)
9. Add the square of these weighted deviations. This gives $S[n_x(\bar{y}_x - \bar{y})^2]$. (This is Σfd^2.)
10. Take the square root of this quantity, and divide by N, the total number of cases in the whole table. This gives

$$\Sigma = \sqrt{\frac{S[n_x(\bar{y}_x - \bar{y})^2]}{N}},$$

the standard deviation of the means of the columns, comparable to

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N}}$$

11. Divide Σ by σ_y, giving *the correlation-ratio, η.*

Since η is the ratio between two standard deviations it is always positive, — that is, η is always between 0 and 1. The expression given here for η is absolutely independent of the *form* of the distribution, whether it exhibits straight-line or curved-line relationship, and can be used for the computation of correlation for any kind of a table. To the writer's knowledge no published analysis has been made of educational distributions which have utilized both the product-moment and the correlation-ratio methods. Thus little comparative data are available for us at this point. We are interested to know: Under what conditions can we use the product-moment formula? How can we determine whether or not a correlation table exhibits *linear regression?* For rough work, Blakeman[1] has stated a criterion for linearity which we can use to aid us with most of our distributions. It is that

$$\frac{\sqrt{N}}{.67449} \cdot \frac{1}{2} \sqrt{\eta^2 - r^2}$$

must be less than 2.5. Applying this to our problem in Diagram 48, we get

$$\frac{\sqrt{148}}{.67449} \cdot \frac{1}{2} \sqrt{(.83)^2 - (-.47)^2} = 6.169 > 2.5.$$

In this case the table is obviously a non-linear table and the product-moment formula is inapplicable. Whenever the correlation table is not very *linear* the investigator should compute both η and r. Then the interpretation of the size of the coefficients η and r can be determined by the application of the criterion for linearity.

B. METHODS WHICH TAKE ACCOUNT ONLY OF POSITION OF THE MEASURES IN THE SERIES

I. Various Methods of Ranks and Grades

From the discussion in the foregoing chapter the two methods of computing correlation which take account of the absolute value and position of each measure in the two

---

[1] Blakeman, J. *Biometrika*, vol. IV, pp. 349, 350.

series have been shown to be mathematically sound, but rather laborious in arithmetical work. It is clear that the student of educational psychology and education often has to content himself with comparatively few subjects, 10 to 30 being quite a common number. With such a small number, the unreliability of the relationships, as shown by the size of the *P.E.*, often would be so great as to vitiate the statistical results.

Other things being equal, that index of correlation is best which gives the smallest *P.E.* With a small number of cases, however, it is clear that the probable error has little or no significance, and that we are unable to establish the reliability of coefficients *computed by any method.*

Spearman's method by " ranks " or " position." At the same time we may desire a practicable formula for the correlation existing between two variables, easily computed and adapted to the conditions of psychological and educational investigation. To supply this formula, Professor C. Spearman had *empirically* deduced a method of expressing correlation in terms of "ranks" or "position," rather than in terms of absolute quantity.[1] This method has been advanced and is coming into common usage, largely on two grounds: (1) the ease of computation of the rank index; (2) the belief that greater comparability of measures will be obtained through expressing the relationships which are found in psychological data in measures of position.

Spearman suggests that a distribution of psychological measures may not be absolutely comparable at various points of the distribution, whereas measures obtained in physical and anthropometrical research may be statistically treated when regarded as being absolutely comparable at all points of the series. On the other hand, Professor William

---

[1] Professor Pearson has since established mathematically the expression for this type of correlation by "grades."

Brown and other psychological pupils of Pearson maintain that the measurement of the results of psychological experimentation *is physical measurement*, and that the measures are objectively comparable.

Spearman has attempted to show that we may turn the product-moment formula

$$r = \frac{\Sigma xy}{N\sigma_x\sigma_y}$$

into the expression

$$r = 1 - \frac{S(v_1 - v_2)^2}{\frac{1}{6}N(N^2-1)}$$

or

$$r = 1 - \frac{6SD^2}{N(N^2-1)},$$

where $(v_1 - v_2)$ or D represents *any difference in the rank* of an individual in the two series, *and where* $\frac{1}{6}N(N^2 - 1)$ *is the value that the sum of the* D^2*'s would have by the operation of chance alone.*[1]

This method is based on a very fundamental assumption, the validity of which is extremely doubtful, — namely, that the distribution of ability is rectangular in shape. This means that "the *unit of rank* is the same throughout the scale," — that is, that individuals are separated from each other at the end of the scale by the same distance (or increment of ability) by which they are separated in the middle of the scale.

Our educational testing of mental abilities leads to the conclusion, however, that most mental abilities distribute themselves in a large school population in accordance with a curve in which the measures are largely concentrated near

---

[1] See Brown, W., *Essentials of Mental Measurement* (1st ed., Appendix), for proof of this statement.

the central portion of the range. It has been shown, *e.g.*, in measuring abilities by various mental tests that the shape of the distributions for each grade and on each test takes a form approximating a symmetrical curve. This curve, with the implications of its widespread use, has been discussed in Chapters VII and VIII. That this relates to the problem of "rank-correlation" should be very clear. To assume a rectangular distribution is to assume that each individual in the series is the same distance from the adjacent individuals, — throughout the series. A glance at the bell-shaped curve shows this to be incorrect. As Pearson says, "Between mediocrities, the unit of rank, . . . is practically zero; between extreme individuals it is very large indeed. Since we must assume a theoretical form of distribution, the form in this case (referring to Spearman's rank-method of computing correlation) must be a rectangle, which is a most improbable one."

Pearson's method by " grades." It has been shown that the best assumption we can make concerning the distribution of ability is that it is somewhat " bell-shaped," that is, *resembles* the normal curve. It happens, therefore, that Pearson has given us a method of computing correlation in which we can use the "grades" (which amount, practically, to "ranks" in actual computation) of each of the measures in the series. There are two points we should clear up, however. (1) The "grade" of a particular individual in a series is measured by the number of individuals above him in the series. (The "rank" indicates the *position* only.) (2) The theoretical distribution of the measures by which the method is worked out is assumed to be that of the "normal" or "probability" curve. This accords more closely with the actual distribution of abilities than Spearman's assumption.

It is possible, therefore, to assume a normal distribution and deduce an expression for r (not Spearman's ρ or R)

measured in terms of the "grade" of an individual in the series.

The expression for the correlation by grades may now be set down as: —

$$\text{(A)} \quad r = 2\ sin\left(\frac{\pi}{6}\rho\right)$$

where

$$\text{(B)} \quad \rho = 1 - \frac{6\ S(v_1 - v_2)^2}{N(N^2-1)} \text{ or } \rho = 1 - \frac{6SD^2}{N(N^2-1)}$$

It will be noted that formula (A) is a sound expression for r, and is unlike Spearman's *empirical* formula for r, which is

$$\text{(C)} \quad r = 2\ sin\left(\frac{\pi}{2}\rho\right)^*$$

In this expression it must be remembered that

$$\frac{1}{6}N\ (N^2-1)$$

is the value that ΣD^2 would be under the operation of chance alone.

The expression

$$r = 2\ sin\left(\frac{\pi}{6}\rho\right)$$

for the correlation of grades, measured in terms of the sum of the squares of the differences of the ranks of all of the measures in the two series, can be shown to be replaced by the following expression when the grades are measured in terms of *the sum of the positive differences* between the grades in the two series.

The formula for r now becomes

---

* For the mathematical development of the theory underlying these expressions the student is referred to the original memoirs by Pearson and his colleagues. (See Appendix.)

$$\text{(D)} \qquad r = 2 \cos 2\pi \left(\frac{\Sigma g}{N^2 - 1} \right)$$

or

$$\text{(E)} \qquad r = 2 \cos \frac{\pi}{3} (1 - R) - 1$$

in which

$$\text{(F)} \qquad R = 1 - \frac{6 \, \Sigma g}{N^2 - 1}$$

We thus have two *complete* formulæ for r, when computed for "*grades*," which are sound mathematically and may be applied, providing the distributions of the traits which are being correlated are approximately "*normal*." These are formulæ (A) and (E) above.

It is clear that the computation by either one may be shortened a great deal by reducing the work as far as possible to the use of tables. It is evident that this can be done for the transmutation of ρ and R into r. Tables VII and VIII (see Appendix) are given herewith for that purpose. Having computed ρ by formula (B), the student can read from Table VII the value of r corresponding to the computed value of ρ. Similarly, for any value of R, the corresponding value of r can be read from Table VIII.

Steps in the computation of r by "rank" methods. Referring to the illustrative problem in Table 42, let us list the steps in the computation of r by these so-called rank-methods.

1. Rank the measures in order of size, beginning with the smallest or largest.
2. Subtract the rank of each measure in the first series from its corresponding rank in the second series. Call this D, the difference in rank. Tabulate these as positive, negative, or 0.
3. If formula (B) is used, square each of these differences, giving the column headed D^2. If Formula (F) is used, treat only the positive differences, the g's of formula (F).
4. Sum the D^2's (or the g's) giving ΣD^2 or Σg.

5. Multiply ΣD^2 or Σg by 6.
6. For formula (B) divide $6\Sigma D^2$ by $N(N^2 - 1)$. N = total number of measures. In the same way for formula (F), divide $6\Sigma g$ by $N^2 - 1$.
7. Subtract the quotient in either case from 1. This is ρ for the first method, R for the second.
8. Transmute ρ into r by reading proper value from Table VII. Transmute R into r by reading proper value from Table VIII.

In the illustrative problem it is noted that $r = .732$ by formula (F), and .717 by formula (B). The conclusion drawn from either one would be the same. In general it may be said that the two formulæ give fairly comparable results, and that from the standpoint of ease of computation the "Footrule" formula

$$R = 1 - \frac{6\Sigma g}{N^2 - 1}$$

may well be the one chosen for use. For small values of N, the only cases after all in which the rank methods are to be used, they lead to as sound conclusions as any of the more accurate methods, the product-moment or correlation-ratio.

Discussion of rank methods of computing correlation. The *first* and principal criticism of Spearman's rank method has been indicated above, namely, that it assumes a rectangular distribution and an equal unit of rank throughout the scale. These assumptions are inadmissible.

Second, Pearson has shown that when the number of cases is small, Spearman's R retains the same value for very wide variations in ρ.

Third, he has shown that the *probable error* of a zero correlation obtained by Spearman's R is considerably larger than that obtained by his r, — hence that "rank" correlations are less accurate than "product-moment" correlations. He says, "In particular it requires about 30 % more obser-

TABLE 42. COMPARISON OF EXPENDITURES PER PUPIL IN AVERAGE DAILY ATTENDANCE FOR VARIOUS SPECIFIC KINDS OF EDUCATIONAL SERVICE.

Computed from the records of the United States Bureau of the Census (Financial Statistics of Cities) and United States Bureau of Education (Annual Report) for the year 1912. To illustrate Computation of Correlation by " rank " methods.*

Salaries of Teachers

City	*Expenditure per pupil*		*Rank in expenditure*		*Difference in rank D*			ΣD^2
	Bureau of the Census	*Bureau of Education*	*Bureau of the Census*	*Bureau of Education*	+	0	—	
Baltimore......	22.43	21.76	15	16	1			1
Boston.........	32.13	29.18	5	7	2			4
Cleveland......	23.50	28.57	14	10			−4	16
Detroit........	28.38	28.91	9	8			−1	1
Jersey City....	25.24	23.96	12	13	1			1
Kansas City...	26.49	25.43	10	12	2			4
Los Angeles....	33.77	41.14	3	1			−2	4
Milwaukee.....	29.91	31.41	8	4			−4	16
Minneapolis....	31.30	31.33	7	5			−2	4
Newark........	20.17	28.32	17	11			−6	36
New Orleans...	22.17	22.90	16	14			−2	4
Philadelphia...	24.07	22.80	13	15	2			4
New York.....	36.15	30.66	1	6	5			25
Pittsburgh.....	31.59	21.03	6	17	11			121
San Francisco..	32.63	32.44	4	3			−1	1
Seattle........	34.32	39.58	2	2		0		
St. Louis.	26.30	28.66	11	9			−2	4
					24		−24	246

$$\rho = 1 - \frac{6\,\Sigma D^2}{N\,(N^2 - 1)} = 1 - \frac{6.246}{17\ (288)} = 1 - \frac{1476}{4896} = 1 - .30 = .70.$$

From Table VII, for $\rho = .70$, $r = .72$.

$$R = 1 - \frac{6\Sigma g}{N^2 - 1} = 1 - \frac{6.24}{288} = 1 - \frac{144}{288} = 1 - .5 = 5.$$

From Table VIII, for $R = .5$, $r = .73$.

Compare $r = .72$ and $r = .73$ obtained by the two methods.

---

* Rugg, H. O. *Public School Costs and Business Management in St. Louis.* (Report of the St. Louis School Survey, 1917.)

vations by the R method to obtain r with the same degree of certainty when r is 0."

Fourth, Spearman's transmutation formula

$$r = Sin(\frac{\pi}{2} . R)$$

was obtained empirically from 111 correlations with only 21 cases ($N = 21$). Brown suggests that the chance that the formula thus selected empirically with but 21 cases was the best one, could not have been great. Many like formulas would have fitted equally well. We should use that formula which has a sound mathematical basis.

In general we may say, that with 30–100 cases or more, that where accuracy is desired in relationships the product-moment method should be used. It gives definite averages (means) and measures of variability, and when tabulated in table form gives a definite perspective of the distribution of measures themselves. In the interpretation of the coefficient it is of great value, — in fact is positively necessary to the adequate interpretation of r. Furthermore, by the use of the correlation table the correlation ratio, η can be computed, which is a necessary step in determining the linearity of regression. Again, ranking the measures introduces a "spurious homogeneity" which may effect the accuracy of our later interpretation and conclusions.

We can thus lay down a rule: USE THE RANK METHOD ONLY WHEN N is small (say, less than 30). In such cases the means and the standard deviations are of little value, owing to the size of the *P.E.*'s. The result in cases of this sort can at best only indicate the *EXISTENCE* of correlation and *Not the Closeness of the Relationship.* Therefore we must be extremely cautious in our interpretation of rank correlations, or of any correlations computed for a small number of cases.

SUMMARY OUTLINE OF METHODS OF DETERMINING RELATIONSHIP

It will pay us, at this point, to summarize in outline form the methods discussed to date, indicating their proper functions: —

I. Methods of Computing Relationship between Series of Measurable Quantities. (Statistics of Variables.)
  1. Methods which take FULL ACCOUNT of the ABSOLUTE VALUE and POSITION of every measure of the series.
    A. *The case of Linear Regression, i.e., the line best "representing" the mean points of the individual columns of the correlation table is a straight line.*

The proper method with N larger than, say, 30 to 50, is the product-moment method

$$r=\frac{\Sigma xy}{N\sigma_x\sigma_y}$$

with the consequent regression equations of the lines of the means of the columns and rows

$$y=r\frac{\sigma_y}{\sigma_x}x, \text{ and } x=r\frac{\sigma_x}{\sigma}y.$$

    B. *The case of Non-Linear Regression, i.e., the case in which the line that best represents the mean points of the correlation table is not approximately a straight line.* The proper method is the "correlation-ratio," η, method of Pearson: —

$$\eta=\frac{\Sigma}{\sigma_y} \text{ or } \eta=\frac{\sqrt{\frac{S[\eta_x(\bar{y}_x-\bar{y})^2]}{N}}}{\sigma y}$$

  2. Methods which take account only of the *position* of measures in the series.
    A. Various methods of Ranks and Grades.
      a. *Ranks.*
        1. Spearman's Method of Rank Differences.

$$\rho = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$$

2. Spearman's "Footrule" for Correlation.

$$R = 1 - \frac{6\Sigma g}{N^2-1}$$

b. Grades. Spearman's Transmutation formulæ are not correct, so we need: —

1. Pearson's Method of Correlation of Grades.

$$(a)\ r = 2\ sin\left(\frac{\pi}{6}\rho\right) \quad \text{in which} \quad \rho = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$$

$$(b)\ r = 2\ cos\ \frac{\pi}{3}(1-R) - 1 \quad \text{in which} \quad R = 1 - \frac{6\Sigma g}{N^2-1}$$

Use Tables VII and VIII (see Appendix) for transmutation to r.

Rough approximation methods. The methods discussed in the foregoing sections have been of two types: (1) refined methods which take full account of *the absolute value and position* of each pair of measures; (2) those which take account *only* of the *rank* or *position* of each pair of measures. There is available to the student, however, a group of rough methods even more approximate in character than the methods of "ranks." These methods take account of *position* of the measures very roughly by classifying the measures with reference to some average point in the two series. We list these methods next, in this outline, prior to discussing them.

B. Various methods of Fourfold Tables.

1. Pearson's:

$$r = cos\ \frac{\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}}\pi$$

2. Sheppard's Method of Unlike-Signed Pairs:

$$r = cos\ \frac{U}{L+U}\pi$$

The methods already discussed in the book have dealt with the statistics of variables, — with problems involving measured quantities of the continuously varying type. It was pointed out in Chapter IV that the student would meet types of problems in which the presence or absence of certain traits would be noted (counted) and in which the correlation methods adapted to statistics of variables would not be applicable. These problems were pointed out under the name "statistics of attributes." Various attempts [1] have been made to devise coefficients which would measure relationships in these types of problems. Most successful of all has been Pearson's coefficient of mean-square-contingency with which we shall close the discussion of relationship. Thus, to complete the outline: —

II. Methods of Computing Relationship between Series of Non-Measured Traits. (The Statistics of Attributes.)
    1. Pearson's Method of Contingency.

B. Methods of Computing Relationship for Fourfold Tables

1. *Pearson's cos π method*

The correlation between the two series of (17) measures in Table 42 was computed by taking account of the relative position, or rank, of each measure in the two series. In this work there was no attempt to measure relative changes in *value* of the measures, except as these were gross enough to change relative ranks. It is evident that a still shorter

---

[1] Yule has devised a "coefficient of association," Q, for fourfold tables. (See Yule, G. U., *An Introduction to the Theory of Statistics*, chaps. II, III, IV, V.)

Pearson, K., and Heron, D. (*Biometrika*, vol. 9, pp. 159–315) have shown that this coëfficient is unstable and rarely leads to sound measures of relationship. Its use is not recommended to the student.

method of computing the extent of relationship could be devised by finding an average of each series of ranks, and comparing the position of each pair of measures with respect to being above or below that average in each series. To do this results in turning the ranking of the two series of measures into a "fourfold table." Tables 43 and 44, and Diagram 50 illustrate this fact.

TABLE 43. RANK OF MEASURES IN TWO SERIES

City	*Rank in first series*	*Rank in second series*
A	1	6
B	2	2
C	3	1
D	4	3
E	5	7
F	6	17
G	7	5
H	8	4
I	9	8
J	10	12
K	11	9
L	12	13
M	13	15
N	14	10
O	15	16
P	16	14
Q	17	11

TABLE 44. RELATIVE POSITIONS OF EACH PAIR OF MEASURES WITH REFERENCE TO AVERAGE OF BOTH SERIES

Above average in both	*Below average in both*	*Above in first; below in second*	*Below in first; above in second*
(*a*)	(*d*)	(*b*)	(*c*)
A			
B			
C			
D			
E			
		F	
G			
H			
I			
	J		
			K
	L		
	M		
	N		
	O		
	P		
	Q		
8	7	1	1

DIAGRAM 50. ILLUSTRATING GROUPING OF MEASURES

	Grouping of measures in the first series	*Grouping of measures in the second series*
Above the median +	A B C D E F G H I*	A B C D E G H I* K
	Median	
Below the median —	J K L M N O P Q	F J L M N O P Q

* With an odd number of cases the middle case must arbitrarily be placed either above or below the median.

Condensing the measures into the number of cases and remembering that in Table 44 —

(*a*) = number of cases above the average in both series,
(*d*) = number of cases below the average in both series,
(*b*) = number of cases above in first series and below in the second series.
(*c*) = number of cases below in first series and above in the second series,

we have: —

$++$ $a = 8$	$+-$ $c = 1$
$-+$ $b = 1$	$--$ $d = 7$

It is clear that such a method of finding correlation takes inadequate account of either position or value of the measures in those cases in which the *form* of the two distributions is not closely the same. For those cases in which the measures are distributed over the scale in approximately the same way, this rough method will supply an adequate measure of correlation, provided a single index can properly be devised for the amount of relationship. Pearson's formula is

$$r = \cos \frac{\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}} \pi\, *$$

Applying this formula to the problem in Table 43, we have

$$r = \cos \frac{\sqrt{1}}{\sqrt{56} + \sqrt{1}} \pi = \cos \frac{1}{8.48} = .118\, \pi$$?

$$= \cos 21.24° = .932$$

---

[*] $\pi = 180°$; the student should have a table of natural trigonometric functions, from which to read the value of r for various values of the angle. This is supplied in the Appendix.

It will be remembered that r by the rank methods gave .717, and .732 by the product-moment method. In general, such approximate methods should be used for only rough preliminary examination.

2. Sheppard's method of unlike signs

Sheppard has suggested an *approximate* formula for *roughly measuring relationship* in fourfold classification in terms of the percentage of cases that are of like or unlike "signs" in the two series of measures.

To get this expression, substitute in Pearson's formula

$$r = cos \frac{\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}$$

for the square root of the product of the bc cases, the percentage of cases having unlike signs (call this U); and for the square root of the ad cases, the percentage of cases having like signs in the two series (call it L). This gives at once Sheppard's formula

$$\text{(N)} \qquad r = cos \frac{U}{L+U} \pi$$

Now, $L + U$ always is 100, and π is 180°. Hence we may reduce the formula to $r = cos\ U\ 1.8°$.

Whipple[1] points out that U must lie between 50 and 0 for positive, and 50 and 100 for inverse correlations, and that therefore it becomes possible to prepare a table from which values of r for any integer of U may be read directly. This table is given herewith as Table IX, Appendix.

The *P.E.* of this

$$r = sin \left(.1686\ \pi\ (1 - r^2) \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \right)$$

---

[1] *American Journal of Psychology*, vol. XVIII, pp. 322–25.

On account of the very large *P.E.* involved in its use, the method of unlike signs must not be used in important correlation work unless the correlations are high (exceeding .50); the classes are very fine, and the number of cases fairly large. Its real function is one of preliminary investigation only. On the other hand, since it involves arranging, in order of size, all the measures in the series, the device is hardly serviceable with large numbers of cases (say 70 to 100 and upwards). For series of 30–50 measures, it might well be used as a method of *preliminary investigation of relationship.*

To illustrate the employment of these methods, Whipple cites an example in which the correlation is desired between the accuracy with which 50 boys can cancel *e* from a printed slip, and the accuracy with which the same 50 boys can cancel *q*, *r*, *s*, and *t* from a similar slip. The results of each test are first arranged in order, the least accurate boy first and the most accurate last. We can either determine the average, in which case all the boys that rank below the average are minus and all that rank above the average are plus, or we can take the median value and consider the first 25 boys in each array as minus and the second 25 as plus cases. The following values were obtained: —

$a = 18$; $b = 11$; $c = 8$; $d = 13$. Hence $U = 38$. By the use of either short formula, $r = .37$ with a *P.E.* of .26. By using Pearson's product-moment method we obtain, for the same arrays, $r = .47$ with *P.E.* of .06. By actual timing, after the distribution had been made, the first method occupied eight minutes and the second two hours and fifteen minutes, even with the adding machine and the tables previously mentioned.

On the other hand, it will be noticed that in the above problem the correlation of .37 with *P.E.* of .26 has absolutely no significance at all, whereas the product-moment value of .47 with *P.E.* of .06 is satisfactory. Furthermore, it should be pointed out that practice in the tabulation of double-entry tables and computation of r by the short method will cut down the time of computation very markedly. Thirty to forty-five minutes should be ample for the computation of r in the above problem.

III. Methods of Measuring Relationship between Series of Attributes

1. *Pearson's coefficient of mean square contingency* (C)

In the foregoing sections methods have been described for treating two kinds of data. The *first* type was data which have been collected in the refined measurement of human traits (known as *Statistics of Variables*). Both refined and approximate methods of treating such measures have been discussed; *i.e.*, detailed regression methods, and approximate rank and fourfold methods. The *second* type of data is that in which we merely count the presence or absence of traits (as when pupils in school pass or fail, are tall or short, are normal or feeble-minded) or in which at the most we classify the data in several groups, without specific quantitative measurement (such as is illustrated by the tables showing relationship between mental age and pedagogical age, in Chapter IV). These kinds of statistics have been called the *Statistics of Attributes*. It is clear that the methods designed to describe relationship between measured quantities are not applicable to the statistics of non-measured traits.

The coarsest method of measuring relationship between such traits is to classify them in a fourfold table, and to treat them by Pearson's or Sheppard's fourfold methods. The weaknesses of these methods already have been pointed out. We need methods which will take cognizance of the classification of measures into several classes and which will be mathematically consistent (as we increase the fineness of classification) with the established theory of the relationship between *variables.*

Pearson's coefficient. Such a method is supplied by Pearson's coefficient of mean-square contingency —

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}}$$

It is built up by reference to the theory of probability, and measures relationship in terms of *the difference between the numbers of measures actually found in the various compartments of the correlation table* (or "contingency" table more generally), *and the numbers that might be expected there by pure chance.*

In Diagram 51 and Table 45 let n_r represent the total number of measures in *any* row of the table, n_c represent the total number in *any* column, N represent the total number in the table, and n_{rc} represent the number in the *compartment* determined by such a row and column. Our first task is to state the *number of measures that ought to fall in any compartment* (say the one determined by the row marked n_r and the column marked n_c) *by pure chance.*

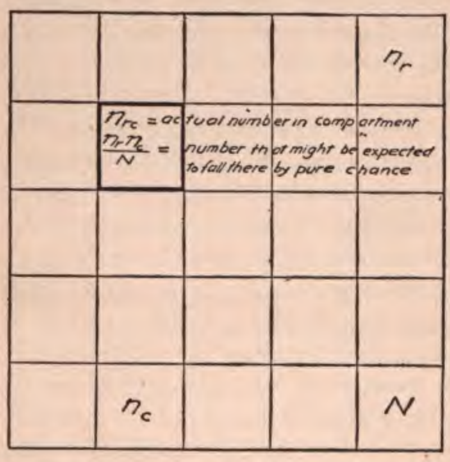


DIAGRAM 51. TO ILLUSTRATE n_c, n_r, n, n_{rc}, AND $\frac{n_r\, n_c}{N}$ IN THE COMPUTATION OF THE "CONTINGENCY COEFFICIENT" C

This can be stated by first stating the probability that any *one* measure will fall in that particular compartment.

Now, the probability that a particular measure will fall anywhere in the row marked n_r is $\frac{n_r}{N}$ and the probability that a particular measure will fall anywhere in the column marked n_c is $\frac{n_c}{N}$. Hence the probability that *any one measure* will fall in both this row and this compartment will be $\frac{n_r n_c}{N^2}$ (the probability of a compound event happening is the product of the probabilities of the separate events.) But we wish the *NUMBER of measures* that ought to fall in this particular compartment. Since there are N measures in the table, this must be N times the probability that any one will fall there. Thus the number that might be expected to fall there by pure chance is

$$\frac{n_r n_c}{N}$$

Since n_{rc} represents the number that actually fell in that compartment, the difference between the two is

$$\Delta = n_{rc} - \frac{n_r n_c}{N}$$

Pearson suggests that a coefficient can be built which will measure relationship by finding the ratios of the differences between the number that actually fall in any compartment and the number that might be expected to fall there by pure chance *to* the number that might be expected to fall there by pure chance. That is by summing the ratios —

$$\frac{n_{rc} - \frac{n_r n_c}{N}}{\frac{n_r n_c}{N}} \qquad (1)$$

We cannot simply add the differences together, for the sum of the values of Δ must be zero (some Δ's are negative, and some are positive), and so we square each of the differences and sum them. If, then, we compute Δ for each compartment, square it, and compute the ratio of each Δ^2 to the corresponding value which is to be expected by pure chance, we can write Pearson's expression for "square-contingency" which will be represented by χ^2, thus: —

$$\chi^2 = \Sigma \left\{ \frac{\left(n_{rc} - \frac{n_r n_c}{N}\right)^2}{\frac{n_r n_c}{N}} \right\} \qquad (2)$$

To give Pearson's *mean-square-contingency*, ϕ^2, we must divide this expression by N —

$$\phi^2 = \frac{1}{N} \Sigma \left\{ \frac{\left(n_{rc} - \frac{n_r n_c}{N}\right)^2}{\frac{n_r n_c}{N}} \right\} \qquad (3)$$

In terms of χ^2 Pearson's coefficient of square-contingency is

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}} \qquad (4)$$

In terms of ϕ^2 his coefficient of *mean-square-contingency* is, since

$$\phi^2 = \frac{\chi^2}{N}, \; C = \sqrt{\frac{\phi^2}{1 + \phi^2}} \qquad (5)$$

It is evident that C is 0 if the two traits are not correlated, and that it approaches more nearly towards unity as χ^2 increases. C is always positive, and no sign should be attached except for conventional purposes.

Yule shows [1] that such coefficients, when "calculated on

---

[1] Yule, G. U., *An Introduction to the Theory of Statistics*, pp. 65 and 66.

different systems of classification, are not comparable with each other. It is clearly desirable, for practical purposes, that two coefficients calculated from the same data, classified in two different ways, should be, at least approximately, identical. With the present coefficient this is not the case: if certain data be classified in, say, (1) 6 × 6-fold, (2) 3 × 3-fold form, the coefficient in the latter form tends to be the least. The greatest possible value is, in fact, only unity if the number of classes be infinitely great; for any finite number of classes the limiting value of C is the smaller the smaller the number of classes."

Yule then shows that Pearson's coefficient of *mean-square-contingency* may be replaced by another which is easier of computation, thus: —

$$\chi^2 = \frac{n_{rc} - \frac{n_r n_c}{N}}{\frac{n_r n_c}{N}}$$

which may be written

$$\chi^2 = \left\{ \frac{(n_{rc})^2}{\frac{n_r n_c}{N}} \right\} - N \qquad (6)$$

For simplicity of statement let the expression

$$\frac{(n_{rc})^2}{\frac{n_r n_c}{N}} \qquad (7)$$

be represented by S.

$$\text{Then } \chi^2 = S - N$$

Then

$$C = \sqrt{\frac{\chi^2}{N+\chi^2}} = \sqrt{\frac{S-N}{N+S-N}} = \sqrt{\frac{S-N}{S}} \qquad (8)$$

This expresses C in terms much easier of computation, and formulas (7) and (8) should be used by the student in computing the relationship between two traits by "contingency."

Yule next shows that if we deal with a $t \times t$-fold classification of data in which the relationship is perfect, "all the frequency is then concentrated in the diagonal compartments of the table, and each contributes N to the sum S. The total value of S is accordingly tN and the value of

$$C = \sqrt{\frac{t-1}{t}} \qquad (9)$$

This is the greatest possible value of C for a symmetrical $t \times t$-fold classification, and therefore, in such a table, for

$t = 2$	C cannot	exceed	0.707
$t = 3$	"	"	0.816
$t = 4$	"	"	0.866
$t = 5$	"	"	0.894
$t = 6$	"	"	0.913
$t = 7$	"	"	0.926
$t = 8$	"	"	0.935
$t = 9$	"	"	0.943
$t = 10$	"	"	0.949

It is well, therefore, to restrict the use of the 'coefficient of contingency' to 5×5-fold or finer classifications. At the same time the classification must not be made too fine, or else the value of the coefficient is largely affected by casual irregularities of no physical significance in the class-frequencies."

Steps in the computation of the coefficient. Taking formula (8)

$$C = \sqrt{\frac{S-N}{S}}$$

we next make clear the steps in the computation of the coefficient. The arithmetic work reduces to four main steps: (1) finding S; (2) subtracting N from S; (3) dividing $S - N$ by S; (4) extracting the square root of $\frac{S-N}{S}$. The detailed procedure is as follows: —

A. Find

$$S = \left(\frac{(n_{rc}{}^2)}{\frac{n_r n_c}{N}} \right)$$

This involves four steps:

(1) Square the number found in each compartment of the table: $(n_{rc})^2$ [*e.g.*, 4, 49, 9, 1, 1, for the first row of Table 45.]

(2) For each compartment in the table multiply the *total* number in *its* column by the total number in *its* row, $(n_r n_c)$ and divide each product by (N), the total number in the table.

For example, for the illustrative problem for the compartments in the *lowest* row: —

$$\frac{2 \times 14}{82} = .34 \qquad \frac{21 \times 14}{82} = 3.59$$

$$\frac{13 \times 14}{82} = 2.22 \qquad \frac{16 \times 14}{82} = 2.73$$

$$\frac{16 \times 14}{82} = 2.73$$

It will probably save time and reduce errors of computation to tabulate these results separately as given by Table 46 below.

(3) For *each* compartment divide the result of doing (1) by the result of doing (2).

For example, for the top row, —

$$\frac{4}{2.83} = 1.41$$

$$\frac{49}{1.48} = 33.11$$

$$\frac{4}{.40} = 10, \text{etc.}$$

(4) Sum each of the results obtained by doing (3). This gives S.

B. Subtract (N), the total frequency of the table, from S, giving $S - N$.

C. Divide $S - N$ by S.

D. Extract square root of $\frac{S-N}{S}$. This gives C, the coefficient of mean-square-contingency.

TABLE 45. RELATION BETWEEN MENTAL AGE AND PEDAGOGICAL AGE

(Computed by coefficient of mean-square-contingency)

		Mental Age in Years							
		9	*10*	*11*	*12*	*13*	*14*	*15*	*Totals*
Pedagogical Age	Retarded 2 years				2		7	2	11
	Retarded 1 year		1		4	9	3	1	18
	Normal			3	8	4	1		16
	Accelerated 1 year		5	10	6	2			23
	Accelerated 2 years		7	3	1	1			14
		2	13	16	21	16	11	3	82

For each compartment compute $\frac{n_r n_c}{N}$, giving the data in the convenient form shown in Table 46.

TABLE 46. DATA GIVING RESULTS OF COMPUTING $\frac{n_r n_c}{N}$ FOR EACH COMPARTMENT OF TABLE 45

		Mental Age						
		9	*10*	*11*	*12*	*13*	*14*	*15*
Ped a gog i cal Age	Retarded 2 years				2.82		1.48	.40
	Retarded 1 year		2.85		4.61	3.51	2.42	.66
	Normal			3.12	4.10	3.12	2.15	
	Accelerated 1 year		3.65	4.49	5.89	4.49		
	Accelerated 2 years	.34	2.22	2.73	3.59	2.73		

These are computed as follows, for the top row: —

$$\frac{21 \times 11}{82} = 2.82 \quad 21 = n_c \quad 11 = n_r \quad 82 = N$$

To compute

$$\frac{(n_{rc})^2}{\frac{n_r n_c}{N}}$$

$^{4}/_{2.82} = 1.42$
$^{49}/_{1.48} = 33.14$
$^{4}/_{.40} = 10$

$^{1}/_{2.85} = 0.351$
$^{16}/_{4.61} = 3.471$
$^{81}/_{3.51} = 23.08$
$^{9}/_{2.42} = 3.727$
$^{1}/_{.66} = 1.515$

$^{9}/_{3.12} = 2.88$
$^{64}/_{4.10} = 15.61$
$^{16}/_{3.12} = 5.13$
$^{1}/_{2.15} = 0.465$

$^{25}/_{3.65} = 6.85$
$^{100}/_{4.49} = 22.27$
$^{36}/_{5.89} = 6.11$
$^{4}/_{4.49} = .891$

$^{4}/_{.34} = 11.735$
$^{49}/_{2.22} = 22.07$
$^{9}/_{2.73} = 3.295$
$^{1}/_{3.59} = .279$
$^{1}/_{2.73} = .367$

$$\begin{aligned} \text{Total} = S &= 174.656 \\ N &= 82. \\ S - N &= 92.656 \end{aligned}$$

$$C = \sqrt{\frac{92.656}{174.656}} = \sqrt{.5305} = .728$$

ILLUSTRATIVE PROBLEMS[1]

1. (*a*) Plot to scale on cross-section paper the following pairs of measures which show the relation between ability of pupils in each of two tests in first-year algebra. Plot Test II on X and Test I on Y. Arrange the work so that this problem (1) and the next problem (2) can be placed on one cross-section sheet.

Test I.....	27	27	27	16	27	18	27	9	15	15	21	20	26	10	22	24	16	13	23
Test II....	20	18	14	3	13	3	16	3	3	7	8	8	17	2	9	20	2	6	9
Test I.....	15	22	20	17	21	20	20	15	22	23	27	16	25	22	14	17	25	22	5
Test II....	2	11	6	8	19	7	9	5	8	16	18	6	12	11	3	7	17	4	2
Test I.....	25	18	27	20	18	27	24	24	24	22	21	20	20	20	24				
Test II....	15	4	23	12	5	22	10	9	12	10	10	13	13	14	13				

(*b*) Plot these same pairs of measures having grouped them in class-intervals of 2 units each.

(*c*) Turn this "point representation" of the pairs of measures into a *correlation-table*, with totals stated on both axes. Use another cross-section sheet for this table, and arrange the work with the tabulation in the upper left-hand corner.

2. (*a*) Plot to scale on cross-section paper the following pairs of measures, which show for United States history the relation between the cost of instruction per 1000 student-hours and the average size of class. Plot the costs on Y and the size of class on X.

Cost........	134	114	26	35	25	62	55	47	46	49	48	55	56	59	51	72	106
Size class....	11	10	38	37	36	23	22	25	24	25	24	22	23	25	24	15	14
Cost........	87	91	114	111	47	53	57	69	35	42	58	31	39	44	105	65	62
Size class....	15	14	12	13	20	21	20	21	27	26	27	29	28	28	17	16	17
Cost........	88	165	137	61	65	72	77	50	38	43	30	40	49	70			
Size class....	12	13	15	19	18	19	18	25	24	30	33	32	33	20			

(*b*) Plot these measures having grouped them in intervals of 2 units.

(*c*) Turn this "point representation" into a "correlation-table." Arrange in upper left-hand corner of the page. Use separate cross-section sheet for (3). Put (1) and (2) on one sheet.

3. Plot the "lines of regression" of the columns and rows for each of the correlation tables plotted in Problems 1 and 2 by the approximate method; (*i.e.*, compute the means of the columns and rows and draw the lines of regression by "cut and try.")

---

[1] Quoted from Rugg, H. O., *Illustrative Problems in Educational Statistics*, published by the author to accompany this text. (University of Chicago, 1917.)

4. Compute the correlation between the following pairs of measures (scores made by pupils in two algebra tests) without tabulation in a correlation-table, by

$$r = \frac{\frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}}{xy}$$

Test I.....	27	27	27	16	27	18	27	9	15	15	21	20	26	10	22	24	16	13	23
Test II....	20	18	14	3	13	3	16	3	3	7	8	8	17	2	9	20	2	6	9
Test I.....	15	22	20	17	21	20	20	15	22	23	27	16	25	22	14	17	25	22	5
Test II.....	2	11	6	8	19	7	9	5	8	16	18	6	12	11	3	7	17	4	2
Test I.....	25	18	27	20	18	27	24	24	24	22									
Test II....	15	4	23	12	5	22	10	9	12	10									

5. For the above data compute the coefficient of correlation by the Spearman "Rank-Coördination" and by the "Foot-Rule" methods.

6. For the above data compute the coefficient of correlation by the "*cos* π" method, and by Sheppard's method of "unlike-signed-pairs."

CHAPTER X

USE OF TABULAR AND GRAPHIC METHODS IN REPORTING SCHOOL FACTS

Studying *vs.* reporting facts. Each chapter of this book has pointed out specific uses of graphic and statistical methods in school practice. Chapter I, especially, gave attention to the use of such methods in the current attempts to solve school problems, by giving typical examples. The use of statistical and graphic methods was shown: in the construction and use of standardized tests; in the preparation of forms for recording school statistics; in the supervision of the teaching of school studies; in the detection of weaknesses in the course of study, and in teaching methods by means of studies in failures; in the comparative method of studying school costs, as shown by Bobbitt's and by Updegraff's early devices; in the use of the probability curve in marking pupils and in standardizing school tests; in the distribution of intelligence in the public schools; and in the use of correlation methods. Throughout the book, the applications have been illustrative of the more refined statistical and graphic methods that can be used in the carrying on of school research, and in the reporting of results to readers technically trained in statistical methods.

The school man, however, having made use of various fairly refined methods in studying his problems, faces the problem of reporting the status of his schools to a public that is, in part, neither trained in the rudiments of statistical method nor familiar with the general conditions of public school administration to-day.

It has been decided, therefore, to conclude the discussion in this book by presenting, in outline form, a representative selection of examples of the application of various tabular and graphic methods of reporting school facts. There is available in print no systematic statement of such methods, brought up to date. Now that school men are beginning to study problems of school administration scientifically, — now that they really are beginning to build up a quantitative knowledge about school conditions, — they are recognizing at last the definite need of ways and means of reporting the facts to the public. School men face no greater problem to-day than that of determining best ways to tell the non-teaching public about the status of schools, and to make clear to them the necessity for doing something about it which will conduce to the definite advancement of school practice.

In reporting school facts to the public we must therefore distinguish the interests and technical equipment of the persons to whom we are reporting our facts. That is, we must recognize that the methods that we should use in reporting experimental and statistical studies to a technically trained group of school people must necessarily differ from the methods with which we should report facts concerning school practice to the general lay citizenship.

The most immediate technical agency (aside from newspapers) for acquainting the public with school conditions is the annual school report. The remainder of this chapter will, therefore, be devoted to a discussion of the form and content of the annual city school report.

School reports are planned and printed to reach three classes of people. These three classes are: (1) administrative officers, teachers, and other school employees, who should be informed of the conduct of school affairs throughout the entire system for the year just finished; (2) professional school officers (interested in either the educational or business aspects of school administration) in other school systems, bureaus, foundations, and professional schools, who are active in studying comparatively the various problems of school administration; and (3) the board of education itself and the more intelligent lay public, whose general insight and educational interest can be depended on to support campaigns for the betterment of the public schools.

Kinds of material that should be included. This clearly must be determined by the aim in mind in attempting to reach the various classes of people to whom the report is to go. It undoubtedly will be agreed that a school report should supply: (1) those facts that can be interpreted and used so as to improve school practice directly, by contributing to the betterment of instruction; (2) those facts that can be interpreted and used so as to improve school practice indirectly, by contributing to the improvement of the work of a non-educational department (buildings or supplies, for example); (3) those facts that will be comprehended by and will stimulate an interest on the part of the general public in the community, and will result in the support of better schools; (4) those facts which will acquaint the public, in accordance with law, with the condition of school property and of school finance in the city.

It can be seen, therefore, that *the criteria of interpretability and of use should largely govern the content of a school report.* The questions to be asked in making up the report should be: Can this statistical table be interpreted so as to improve some phase of school practice? Does it provide comparative material of which other school systems or students of school administration can make use? Can these data be understood by the public, and has the interpretation and explanation been made so complete and clear that this report will operate as a means of "educating" — as well as informing — the public to active support of the public schools?

Again it undoubtedly will be agreed that a school report should contain material of three distinct types: (1) *Current material:* It should report the local situation in sufficient detail to explain the significant developments of the current year and the present condition of the public schools. (2) *Historical material:* It should present enough historical statistics concerning the growth of important phases of the schools' work to permit a discussion of particular aspects and of relative efficiency of the activity of certain departments. (3) *Comparative material:* It should contain comparative data of the procedure of other school systems working under similar conditions. Lacking an absolute standard for judging the efficiency of school practice, the common practice of many cities may well serve as a check upon the methods employed in any one.

Statistical material must be interpreted by descriptive material. To include pages of statistical material with no interpretations or comparisons is, for the layman at least, a waste of printer's ink. All tabular and graphic data should be interpreted clearly, either by the officer who publishes the material, by the superintendent of the schools, or by some other officer especially appointed in the system to study ways and means of improving the conduct of school business, — for example, the director of the "bureau of school research and efficiency." Thus, school reports which have been very largely "statistical" and "informational" should become "educational" in the widest community sense. The school report in a city system can be made a valuable instrument for the promotion of school work in the city. To become that, however, it at least must conform to the foregoing fundamental criteria.

Important criteria concerning the form of the school

report. We may discuss briefly the more important questions arising in connection with the form of the annual school report. The first question to be settled is this: Shall the school report be one volume appearing annually, biennially, or less frequently, or shall it be published in the form of a series of short monographs, each of which discusses one phase of school work? The traditional school report is a composite volume made up of general descriptive articles by educational officers of the system, put together in one portion of the report, and followed by a large mass of statistical data, as a rule completely uninterpreted and, on the whole, uninterpretable by the general public. Such a volume, inquiry has shown, is almost never read by any portion of the public.

A great advance has been marked out by the recent innovation begun by Superintendent F. E. Spaulding, while he was Superintendent of Schools at Minneapolis, Minnesota, in the publication of a series of monographs describing in clear language, and pertinently illustrated by graphic and comparative statistical methods, the status of educational activities of the city of Minneapolis. In the quotations of this chapter, we shall make several references to this excellent practice.

We cannot decide the question of the general form of the school report, however, without taking account of the question of the frequency with which school facts need to be reported. Should all data regarding school practice be reported annually, or are there types of facts which may well be published but intermittently?

Classes of school facts. We can distinguish school facts, therefore, in two classes: —

First, those that are reported annually. These may be summarized as follows: (1) facts that either state or municipal law requires must be published each year, concerning the extent and condition of school property; (2) current and historical local statistics concerning the financial condition of the board of education, the distribution of pupils according to ages and grades, the enumeration of children of school-census age in the city, and the detailed reporting of statistics on the teaching staff; (3) facts concerning the progress of educational experiments or innovations that have been established prior to the current year, and in which the public will have a definite interest: for example, new methods of detecting defects in children, and of providing for them; special forms of instruction; new developments in vocational education; etc.; (4) information concerning the establishment of new educational experiments, — important and far-reaching changes in the administration of the local schools, etc.

Second, school facts that are reported intermittently. It frequently occurs that it is necessary for the superintendent and the board of education to give the public detailed and specific information concerning needed enlargements, greater financial support of the schools, etc. For example, our larger cities are all feeling the need for increased revenues for permanent improvements to the school plant. School populations are increasing, and the consequent demands on our public school facilities are likewise increasing, usually more rapidly than are the revenues made possible under state law by the increase in real wealth in the community. City school boards are finding it imperative, therefore, to go to the people for authority to bond the school district in order to finance the additional school plant which is needed. This necessitates an educational campaign, and this in turn demands a special kind of school report. This report may well give facts to the public that ordinarily will not need to be given each year. For example, a detailed comparative analysis of the status of school finance in this particular city with that in other comparable cities, together with an analytical study of the historical development of various aspects of school finance may well be needed. We shall point out, later on, illustrative methods of reporting such studies.

School facts that should not be printed at all. Careful examination of current city school reports reveals the publication of many types of statistical and descriptive material that ought not to be printed at all. This can be illustrated partially by listing specific types of non-usable statistics published in the annual report of one of our largest and most progressive school systems: (1) tables of total values of supplies delivered to various types of schools (of little value unless reduced to some unit basis, and presented historically); (2) analyzed statement of total cost of transportation by schools for current year; (3) a table, twelve pages long, giving itemized amounts of each particular kind of supplies delivered to each building in the system; (4) list of textbooks lost or destroyed in district schools, giving names of the books, number, and price of each; (5) number and money value of condemned books, together with rebound books, by specific title, number, price, etc.; (6) list of textbooks, giving name of book, number in usable condition in all public schools, price, value, etc. (16 pages); (7) names of pupils graduating from various schools; (8) names and facts concerning all teachers and other officers on the staff; (9) detailed statement of total expenditures for particular activities for each building in the system (as "totals" the table is uninterpretable; it might be condensed to small fraction of present size, 44 pages; it ought to be reduced to a per-pupil basis); (10) detailed statements concerning cost of particular activities and special schools, giving totals and itemized expenditures, etc., — might well be condensed and published as "unit" costs.

I. Content of the Annual School Report: Suggestive Examples of Tabular and Graphic Methods

The foregoing introductory discussion can now more immediately be focused upon the specific organization of the content of the school report. As we proceed with the discussion, in each case we shall point out whether the material should be annually reported or reported at intervals of several years.

1. Legal basis of the local school system

Form of organization. The introductory statements should contain a table of contents, with a pertinent list of subheadings, to make clear to the reader the important points discussed in the report. This should be followed by a brief text statement describing the legal basis of the system. The reader should be told the important facts concerning the origin, development, and present legal status of the city school district, exactly how its functions are affected by those of city civil district, and what important changes have come about in this legal status. A clear statement should be given concerning the present board of education — its size, how members are selected, the specific powers and duties of the board, the committee organization, tenure, compensation of board members, etc.

Legal basis of school finance. This should be pointed out very clearly, answering such questions as: Does the board of education have complete tax-levying power? If not, by what agency are its budgets reviewed? What are the legal limits of school revenue? Are permanent improvements and current school expenses financed out of taxation? What is the legal status of bonding the school district for school purposes, and of borrowing for temporary purposes on short-term notes?

The detail into which the annual discussion of the legal basis of school finance should go must be determined by the financial condition and by the current financial powers of the board. A brief statement of the latter is all that is required in an annual report in which special efforts are not being made to effect a change in taxing powers, taxing limits, etc. In case it becomes necessary to make a special plea to the people, the report should go into the legal status carefully. If the critical change needed is to give the board of education complete taxing power, and the board wishes to show the effects of having its budgets reviewed by another governmental body, a table such as Table 47 and a diagram such as Diagram 52 might be used, with proper textual explanations.

TABLE 47. COMPARISON OF THE BOARD OF EDUCATION AND COMMON COUNCIL BUDGETS OF GRAND RAPIDS, MICHIGAN, TOGETHER WITH AMOUNTS SPENT FOR PERMANENT IMPROVEMENTS, 1910–11 TO 1915–16 INCLUSIVE.

(*Data from Official Proceedings of the Board of Education*)

Year	*Board of education budget*	*Common council budget*	*Total amount spent for permanent improvements*	*Amount included in common council budget to be devoted to payment of bonds and interest*
1910–11	$201,443.79	$107,897.11	$404,466.14	$49,860
1911–12	183,166.50	121,166.50	245,751.97	78,792
1912–13	103,785.50	97,055.50	157,159.14	80,577
1913–14	100,089.00	100,089.00	89,880.59	64,095
1914–15	273,792.00	126,792.00	249,594.73	101,292
1915–16	233,310.00	98,960.00	545,771.48	77,960

2. Presentation of facts concerning school revenues and expenditures

The superintendent of schools annually will wish to make clear to his community the following facts concerning school finance: —

(*a*) **Comparison of total possible school revenue and actual school revenue.** This would mean the total possible tax levies for school purposes (computed from the assessed property valuation and the legal limit, in mills on the dollar

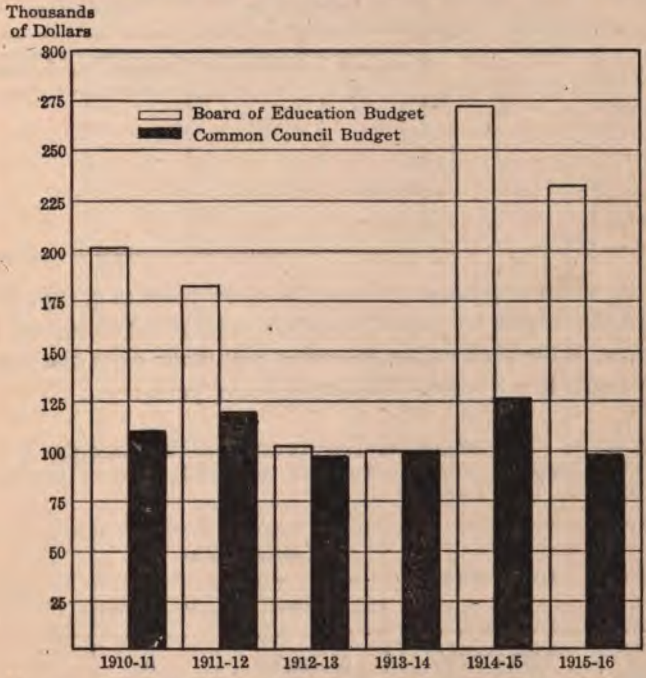


Diagram 52. Comparison of Board of Education Budget for Permanent Improvements with Budget approved by Common Council

of assessed valuation), compared with the actual tax levy for school purposes, and covering a series of years. Throughout the entire school report the presentation should distinguish definitely between school finance for *current purposes* and for *permanent improvements.* To get the situation clearly before the reader, diagrams such as Diagrams 53 and 54 can be used effectively. If tables are to be given, the headings and data can be organized somewhat as follows: —

Year	*Assessed property valuation*	*No. of mills*	*Possible tax levy*		*Actual tax levy*	
			For current purposes	*For permanent improvements*	*For current purposes*	*For permanent improvements*
1906						
1907						

1915						

In discussing the relation between school-taxing capacity and the degree to which the city is taking advantage of it, a brief table such as the following will make clear the most probable taxative possibilities in future years: —

TABLE 48. COMPARISON OF ESTIMATED POSSIBLE SCHOOL TAXING CAPACITY FOR YEARS 1920 TO 1930, WITH PROBABLE ACTUAL TAX LEVIES *

Year	*Assessed valuation*	*Actual school tax*		*No. of mills possible*
		Amount	*No. of mills*	
1920	$203,000,000	$1,001,000	4.98	6
1925	243,000,000	1,276,000	5.26	6
1930	283,000,000	1,551,000	5.48	6

* Example quoted from Rugg, H. O., *Cost of Public Education in Grand Rapids*, p. 369. (1917.)

(*b*) Sources and amounts of revenue. A table should be printed each year that will present concisely the sources and amounts of revenue during a series of recent years, classified under such headings as: (1) balance on hand; (2) received

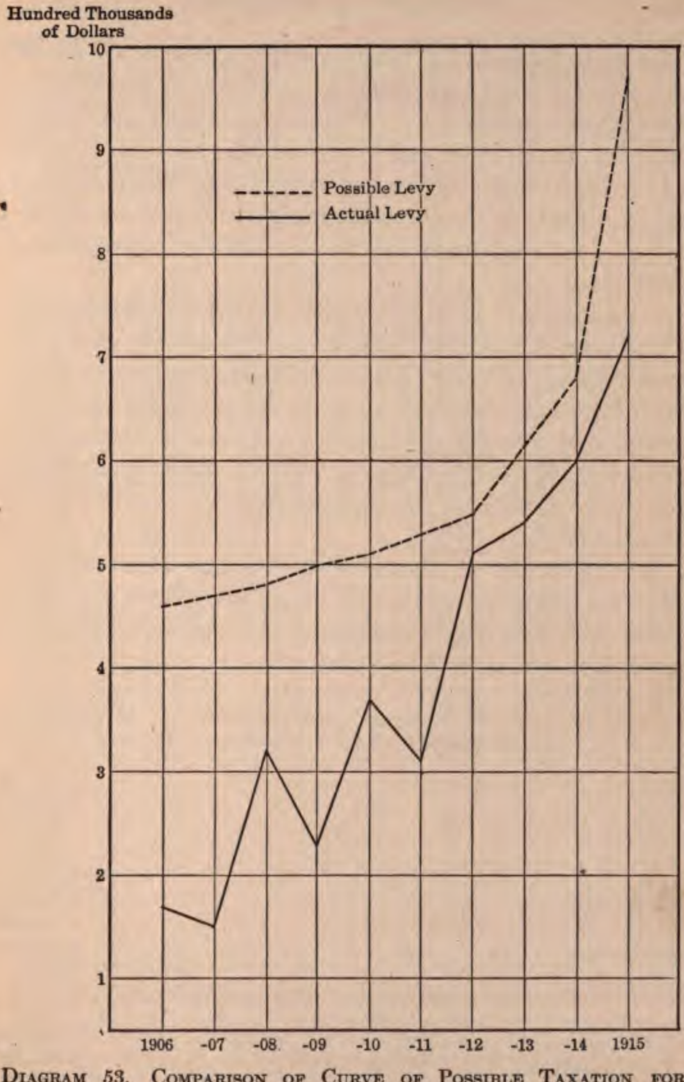


DIAGRAM 53. COMPARISON OF CURVE OF POSSIBLE TAXATION FOR GENERAL PURPOSES WITH ACTUAL TAX LEVY, 1906–15

from state sources; (3) from county sources; (4) from city sources; (5) from miscellaneous sources; (6) total income for annual maintenance; (7) from sale of bonds; and (8) total receipts. It might be well to add another short table giving percentages that each source contributed to the total receipts. Each of these items should be shown for a series of years, at least ten, so as to admit of comparisons being easily made.

(*c*) **Relation of revenue receipts to current expenditures.** Each year it would be well to publish a table giving the facts for a series of years relating to receipts and expenditures, both for current maintenance and for permanent improvements, and showing the financial condition of the board of education by comparing the receipts with expenditures for each and showing the surplus or deficit each year for a number of years.

(*d*) **Methods of financing permanent improvements.** If a table showing the source of all receipts for a series of years and such charts as Diagrams 53 and 54 have been presented, the degree to which the city is paying for its

TABLE 49. SCHOOL BONDED INDEBTEDNESS IN MINNEAPOLIS ADDED DURING THE LAST FIVE YEARS

1911	$1,116,700
1912	500,000
1913	775,300
1914	825,000
1915	1,125,000
Total for five years	4,342,000
Bonds redemmed in same five years	80,000
Net increase	4,262,000

SCHOOL BOND ISSUES FROM 1889 TO 1916

School bonds outstanding December 31, 1889	$542,500
School bonds issued, 1889–1916	6,825,000
School bonds redeemed during 26 years	292,500
School bonds outstanding December 31, 1915	7,075,000

TABLE 50. MINNEAPOLIS BONDED SCHOOL DEBT COMPARED WITH SCHOOL DEBTS OF OTHER CITIES, DECEMBER 31, 1915

Cities of 200,000 population or more — 1915 estimate	*Amount of school bonds outstanding*	*Per capita school debt*	*Total net debt*	*Per capita total debt*	*Ratio of school debt to total debt —per cent*	*Ratio of school debt to value of school property—per cent*
New York.....5,468,190	$123,425,000	$22.57	$991,219,000	$181.27	12.45	....
Chicago.......2,447,045	....	....	39,423,000	16.11	....	....
Philadelphia...1,683,664	13,827,000	8.23	104,823,000	62.25	13.19	....
St. Louis...... 745,938	....	....	19,579,000	26.24	....	....
Boston........ 745,139	16,227,000	21.78	84,423,000	113.34	19.22	....
Cleveland..... 656,905	6,946,000	10.57	56,242.000	85.61	12.35	....
Baltimore..... 584,685	3,300,000	5.54	67,064,000	114.64	4.96	....
Pittsburgh.... 571,914	10,703,000	18.71	42,923,000	75.04	24.93	58.53
Detroit 554,766	8,767,000	15.80	17,563,000	31.64	49.91	66.45
Los Angeles... 475,337	8,635,000	18.18	45.696,000	96.20	18.89	65.63
Buffalo........ *461,305	7,411,000	16.07	38,095,000	82.64	19.46	70.96
San Francisco. 448,502	5,779,000	12.90	42,172,000	93.92	13.70	....
Milwaukee.... 428,062	3,537,000	8.27	11,921,000	27.85	29.66	....
Cincinnati 406,706	4,588,000	11.27	61,170,000	150.30	7.48	....
Newark....... 399,000	8,922,000	22.36	30,864,000	77.35	28.91	86.15
New Orleans.. 366,484	....	....	37,088,000	101.33	....	....
Washington... 358,679	Bonds not issued for specific purposes		6,287,000	17.56	....	....
Minneapolis.. 353,460	**7,075,000**	**20.04**	**19,906,000**	**56.39**	**35.54**	**85.45**
Seattle........ 330,834	4,750,000	14.35	21,807,000	65.88	21.78	....
Jersey City.... 300,133	4,492,000	14.97	19,397,000	64.66	23.16	....
Kansas City... 289,879	7,823,000	26.98	10,733,000	37.00	72.89	80.51
Portland...... 272,833	849,000	3.11	15,980,000	58.53	5.31	....
Indianapolis .. 265,578	2,007,000	7.55	6,369,00	23.94	31.52	43.06
Denver 253,161	....	....	2,946,000	11.64	....	....
Rochester..... 250,747	1,428,600	5.69	17,996,000	71.70	7.94	33.49
Providence.... 250,025	2,697,000	10.79	11,138,000	44.55	24.22	62.28
St. Paul 241,999	1,990,000	8.22	11,359,000	46.94	17.52	50.90
Louisville..... 237,012	937,300	4.08	11,995,000	50.61	7.81	30.33
Columbus 209,722	1,457,200	6.94	11,260,000	53.62	12.94	....
General average of the per capitas...........	....	10.85	....	66.48	....	....
Average of percentages	....	....	....	....	21.41	61.14

* Estimate, 1914.

school property "as it goes" will have been shown. The policy of the city in the use of school bonds, and the condition of school and city indebtedness should be shown, especially if a special campaign is being carried on for funds. Tables 49 and 50 show how Superintendent Spaulding pre-

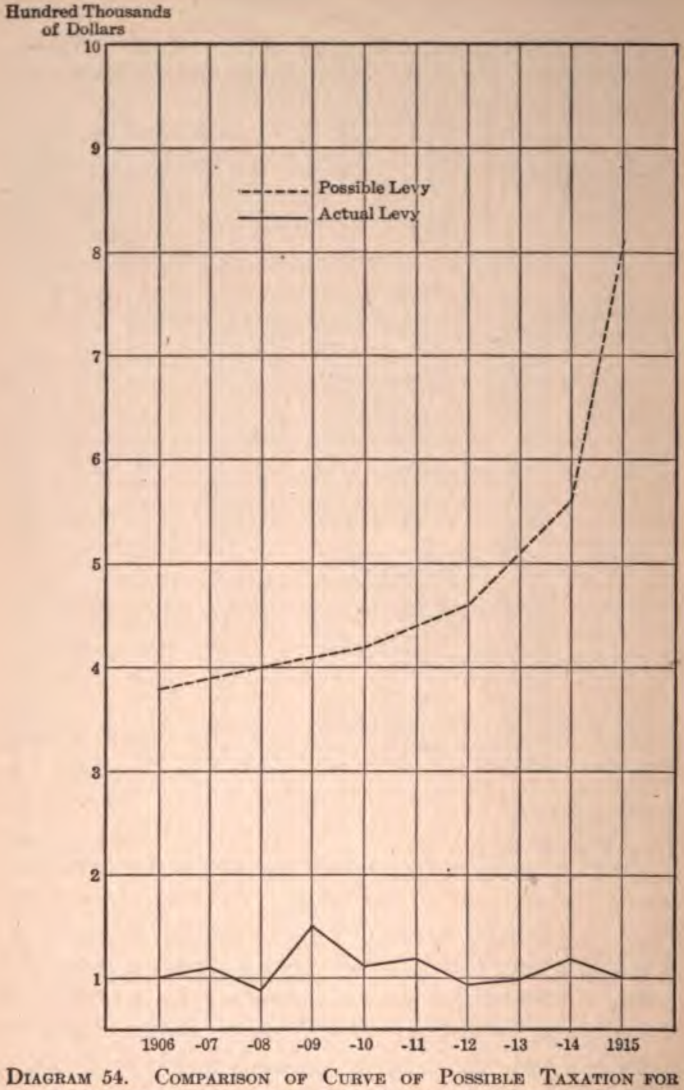


DIAGRAM 54. COMPARISON OF CURVE OF POSSIBLE TAXATION FOR PERMANENT IMPROVEMENTS WITH ACTUAL TAX LEVY, 1906–15

sented the data in one of his 1916–17 School Monographs.[1] Columns might well have been added giving the rank of the cities in question.

Diagram 55. Total City and School Bonded Indebtedness, 1890–1915

It is probable that a line diagram of the nature of Diagram 55 will do much to clarify such a presentation. If possible the data ought to be given for each year, as in that diagram. The present indebtedness of the school city can be cleared up further by a table giving the total outstanding bonds maturing each year. Table 51 suggests the form.[2]

---

[1] For excellent descriptive and graphic methods of reporting such facts see three monographs published by the Minneapolis Board of Education: *Financing the Public Schools; A Million A Year; The Price of Progress,* 25c each.

[2] From *Annual Report of Business Manager* (1915), Grand Rapids, Michigan.

TABLE 51. TOTAL AMOUNTS OF OUTSTANDING BONDS MATURING EACH YEAR, 1916 TO 1930

(Data from 1915 Report of the Board of Education)

Year — June 30 to July 1	*Principal*	*Interest*	*Total*
1915–16	$35,000.00	$41,935.00	$76,935.00
1916–17	63,000.00	39,912.50	102,912.50
1917–18	75,000.00	37,002.50	112,002.50
1918–19	75,000.00	33,727.50	108,727.50
1919–20	75,000.00	30,352.50	105,352.50
1920–21	75,000.00	26,977.50	101,977.50
1921–22	75,000.00	23,602.50	98,602.50
1922–23	100,000.00	19,752.50	119,752.50
1923–24	50,000.00	16,490.00	66,490.00
1924–25	75,000.00	13,702.50	88,702.50
1925–26	70,000.00	10,440.00	80,440.00
1926–27	....	8,865.00	8,865.00
1927–28	64,000.00	7,425.00	71,425.00
1928–29	75,000.00	4,297.50	79,297.50
1929–30	58,000.00	1,305.50	59,305.50
Total outstanding.	$965,000.00	$315,788.00	$1,280,788.00

(*e*) **Capacity of the city to support schools, and degree to which it is doing so.** This can be shown by stating the city expenditures, — first, per inhabitant, second, per $1000 of real wealth in the city, and, third, per pupil in average daily attendance. This further calls up the question of comparing expenditures in the local city with those in a group of comparable cities. How often should such a comparative analysis be made? It is probable that lack of clerical assistance will prevent the compilation each year of original data, and the computation of unit costs with consequent "ranking" of cities. It certainly should be done every few years. If it can be done, Tables 52 and 53[1] and Diagram 56[2] suggest the type of comparison that can be made to establish the point at hand.

---

[1] Clark, E., *Financing the Public Schools*, pp. 27 and 29. [2] *Ibid.*, p. 33.

TABLE 52. EXPENDITURE PER INHABITANT FOR OPERATION AND MAINTENANCE OF SCHOOLS IN CLEVELAND, AND IN 17 OTHER CITIES OF FROM 250,000 TO 750,000 INHABITANTS, 1914

City	*Estimated popula- in 1914*	*Expenditure for operation and maintenance*		*Rank in expenditure per inhabitant*
		Total	*Per inhabitant*	
Baltimore......	579,590	$1,954,670	$3.37	17
Boston.........	733,802	5,516,762	7.52	2
Buffalo.........	454,112	2,449,533	5.39	12
Cleveland.....	**639,431**	**3,569,504**	**5.58**	**9**
Detroit........	537,650	2,553,488	4.75	14
Indianapolis ...	259,413	1,409,504	5.43	11
Jersey City....	293,921	1,421,147	4.84	13
Kansas City...	281,911	1,761,389	6.25	7
Los Angeles...	438,914	3,706,519	8.45	1
Milwaukee	417,054	1,794,796	4.30	15
Minneapolis ...	343,466	2,147,856	6.25	6
Newark.......	389,106	2,699,239	6.94	3
New Orleans..	361,221	1,097,552	3.04	18
Pittsburgh.....	564,878	3,602,303	6.38	5
San Francisco..	448,502	1,879,187	4.19	16
Seattle........	313,029	1,750,988	5.59	8
St. Louis......	734,667	4,084,693	5.56	10
Washington....	353,378	2.391,976	6.77	4
Average.....	....	....	$5.59	....

Other graphic methods. The four diagrams which follow show means which may be used by superintendents to reveal facts to their constituencies, using graphic instead of tabular methods of presentation. A little thought given to devising such graphic representations at the time of preparing the annual report will be time well spent.

Chapter II discusses in detail the sources and validity of such comparative school statistics. Cities should always be

Table 53. Expenditure per $1000 of Wealth for Operation and Maintenance of Schools in Cleveland, and in 17 Other Cities of from 250,000 to 750,000 Inhabitants, 1914

City	*Estimated true value of all property assessed*	*Expenditure for operation and maintenance*		*Rank in expenditure per $1000 of property assessed*
		Total	*Per $1000 of property assessed*	
Baltimore	$723,800,340	$1,954,670	$2.70	17
Boston........	1,489,608,820	5,516,762	3.70	11
Buffalo........	494,200,459	2,449,533	4.96	3
Cleveland.....	**756,831,185**	**3,569,504**	**4.72**	**5**
Detroit........	598,634,198	2,553,488	4.27	9
Indianapolis ...	363,413,650	1,409,504	3.88	10
Jersey City....	257,644,605	1,421,147	5.52	2
Kansas City...	371,191,014	1,761,389	4.75	4
Los Angeles....	836,604,260	3,706,519	4.43	8
Milwaukee.....	511,720,797	1,794,796	3.51	14
Minneapolis....	639,258,841	2,147,856	3.36	16
Newark.......	383,864,182	2,699,239	7.03	1
New Orleans...	314,086,036	1,097,552	3.49	15
Pittsburgh	789,035,200	3,602,303	4.57	6
San Francisco..	1,247,391,284	1,879,187	1.51	18
Seattle.	473,174,995	1,750,998	3.70	12
St. Louis......	1,125,308,749	4,084,693	3.63	13
Washington....	538,389,607	2,391,976	4.44	7
Average.....	....	....	$4.12	..

selected for ranking purposes which are comparable as to (1) population, (2) geographical location, (3) wealth, and (4) legal status.

(*f*) **Extent to which city supports schools as compared with way in which it supports other city departments.** This can be presented if careful study shows the necessity. The data can be found in an annual publication of the United States

Bureau of the Census (*Financial Statistics of Cities*). If the data are used, three comparative tables should be given stating: (1) amount spent per inhabitant for various city

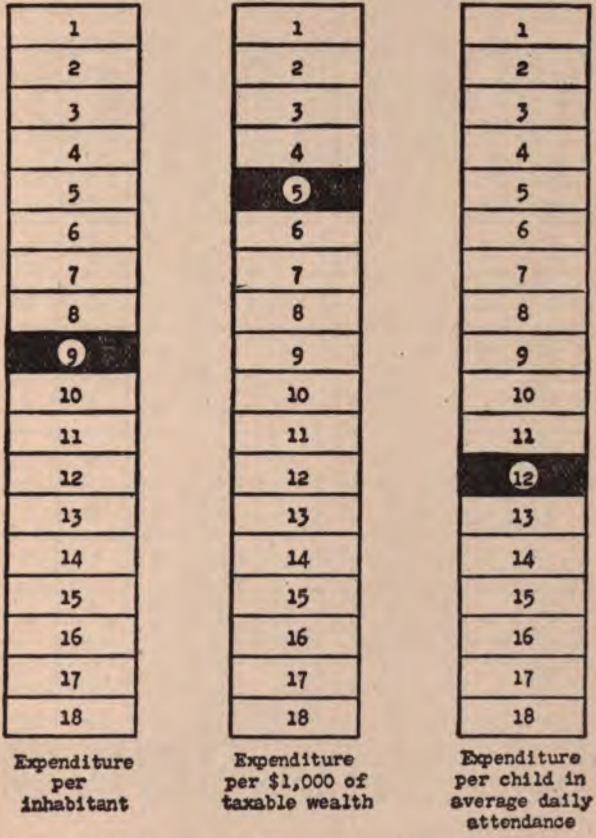
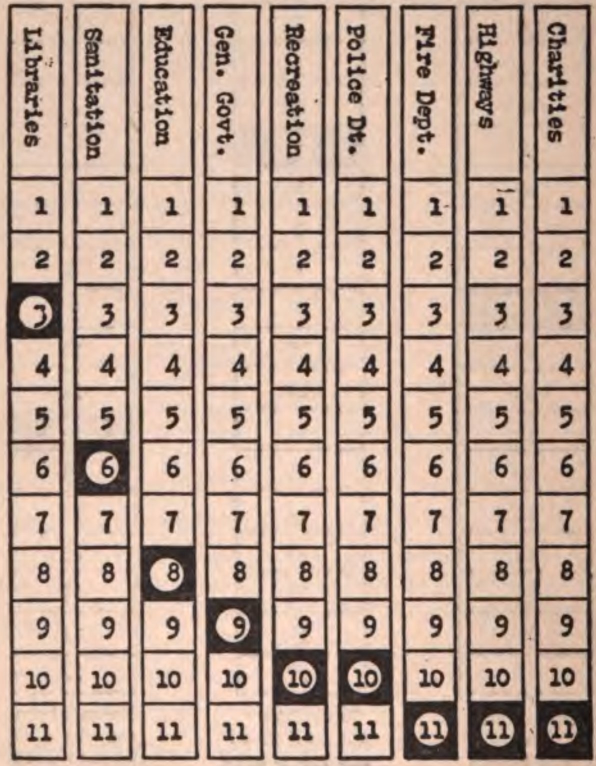


DIAGRAM 56. RANK OF CLEVELAND IN GROUP OF EIGHTEEN CITIES IN EXPENDITURE FOR OPERATION AND MAINTENANCE OF SCHOOLS

Given per inhabitant, per $1000 of taxable wealth, and per child in average daily attendance. (From Ayres, L. P., *The Cleveland School Survey*, 1916.)

departments, including schools; (2) per cent of total governmental cost payments devoted to various city departments; and (3) rank in per cent of total governmental cost payments devoted to various city departments. Diagram 57 is re-



Libraries	Sanitation	Education	Gen. Govt.	Recreation	Police Dt.	Fire Dept.	Highways	Charities
1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2
③	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5
6	⑥	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7
8	8	⑧	8	8	8	8	8	8
9	9	9	⑨	9	9	9	9	9
10	10	10	10	⑩	⑩	10	10	10
11	11	11	11	11	11	⑪	⑪	⑪

Diagram 57. Rank of Cleveland among Eleven Large Cities in per capita Expenditures for each Principal Kind of Municipal Activity

Numbers in black circles show Cleveland's rank. (From Ayres, L. P., *The Cleveland School Survey*, p. 26.)

produced [1] to illustrate the departments considered and the method.

(*g*) **How the board of education spends its money.** The publication of current total and per capita expenditures are of little value, unless they are accompanied by historical and comparative statistics. The school report should give: —

First. The total amounts spent, and the amounts spent per pupil in average daily attendance, for (1) all current expenses, and (2) permanent improvements. This can be pictured clearly by a chart after the form of Diagram 53.

Second. The total and per pupil (in average daily attendance) [2] expenditures for all educational purposes, as contrasted with all business purposes and the per cent devoted to each. Ranks of the cities should be given for each table. A five-year table giving the relative expenditures in the local city might well be included.

Third. The degree to which the board supports different kinds of educational service: (1) for the larger aspects, such as administration, supervision and instruction, operation of plant and maintenance of plant; and (2) for specific kinds of service, such as board of education office, superintendent's office, salaries of supervisors and their clerks, salaries of principals and their clerks, salaries of teachers, stationery and educational supplies, wages of janitors, fuel, water, light and power, and repairs. For each of these items the reporting should be done in terms of (1) total amount spent; (2) amount spent per pupil; (3) per cent of total expenditures devoted to each; and (4) rank of all cities in the list, for the expenditures for each item, in order to compare the local city with other cities of its class.[3]

---

[1] Ayres, L. P., *The Cleveland School Survey*, p. 26.

[2] For items to include under each see Clark, E., *Financing the Public Schools*, p. 65; or *Grand Rapids School Survey*, p. 388. Detailed tables and forms are given in the latter.

[3] For suggestions see Clark, E., *Financing the Public Schools;* or Rugg, H. O., *Cost of Public Education in Grand Rapids*.

Fourth. Total expenditures and expenditures per pupil for capital outlay. Because of the fluctuations from year to year in expenditures for permanent improvements, such ought to be reported both for the current year, and for an average of four or five years. If clerical assistance makes it possible, this should also be compared with the other cities in the list. It involves very laborious computations, if done for many years.

Fifth. The degree to which the board supports different kinds of schools, — elementary and secondary. Table 54 suggests a comparative method of reporting this aspect of

TABLE 54. DISTRIBUTION OF CURRENT EXPENDITURES FOR ELEMENTARY AND SECONDARY SCHOOLS — 17 CITIES, 1915 *

(*Data from United States Commissioner's Report, 1915, vol. 2*)

City	*Per cent of total current expenditures devoted to*		*Rank in per cent of total current expenditures devoted to*		*Expenditure per pupil in average daily attendance*		*Rank of 17 cities in expenditure per pupil in average daily attendance for*	
	Elementary schools	*Secondary schools*	*Elementary schools*	*Secondary schools*	*Elementary schools*	*Secondary schools*	*Elementary schools*	*Secondary schools*
Albany	77.44	22.56	11	7	35.69	70.56	4	5
Birmingham	81.78	18.22	7	11	23.71	49.10	16	16
Bridgeport	88.39	11.61	1	17	26.01	54.95	14	12
Cambridge	75.29	24.71	14	4	36.35	63.58	3	9
Dayton	72.37	27.63	17	1	29.85	63.77	10	8
Des Moines	77.25	22.75	12	6	33.66	51.17	6	14
Fall River	79.86	20.14	10	8	34.92	87.32	5	3
Grand Rapids	76.65	23.35	13	5	40.45	87.36	2	2
Lowell	82.08	17.92	5	13	31.37	47.27	9	17
Lynn	74.72	25.28	15	3	27.49	65.42	13	7
Nashville	82.25	17.75	4	14	24.37	56.57	15	11
New Bedford	82.07	17.93	6	12	32.46	84.02	7	4
Paterson	82.82	17.18	3	15	27.65	51.88	12	13
Richmond	80.10	19.90	9	9	22.24	56.73	17	10
San Antonio	83.62	16.38	2	16	31.51	50.66	8	15
Scranton	81.26	19.73	8	10	27.75	66.70	11	6
Springfield	73.52	26.48	16	2	44.64	94.74	1	1

---

* Rugg, H. O., *Cost of Public Education in Grand Rapids.* (1917.)

school finance to the public. It presupposes the publication of the total expenditures for elementary and secondary schools, together with the per cent of all expenditures devoted to each, and the unit expenditure per pupil in average daily attendance. Ranks of all cities are given for both sets of data.

3. *The reporting of facts concerning the teaching staff*

Data to be reported. The numerical status of the city's teaching staff, in each of its various departments, should be reported each year. It may be presented compactly, together with various historical data, in a table such as Table 55. Line diagrams of the sort shown for enrollment in Diagram 60 may well be drawn to picture the status more clearly.

It is desirable to present the facts on the distribution of the teaching staff according to ranks and salaries, as com-

TABLE 55. DISTRIBUTION OF SCHOOL OFFICERS AND TEACHERS IN DIFFERENT GRADES OF SCHOOLS, 1910–1915 INCLUSIVE*

Year	*High school principals and assistants*		*Elementary schools*			*Kindergarten*			*Manual training*			
Sept.	*Number of teachers*	*Principals*	*Principals*	*Regular teachers*	*Ungraded teachers*	*Teachers*	*Assistants*	*Supervisors*	*Supervisors or directors*	*Teachers*	*Auxiliary teachers*	*Special supervisors*
1910..	64	4	34	295	10	1	35	16	1	25	4	3
1911..	77	4	34	300	16	1	35	17½	1	28	6	2
1912..	79	2⅔+	33⅓	314	21	1	35	20	1	18	6	6
1913..	81	3⅔	34⅓	329	23	1	35	26	1	25	6	7
1914..	92	3⅔	36⅔	334⅔	20	1	36	30	1	26	8	8½
1915..	114	4⅓	36⅙	349½	18½	1½	34½	31	2	33	12	14½

* Rugg, H. O., *Cost of Public Education in Grand Rapids.*

TABLE 56. SHOWING DISTRIBUTION OF TEACHERS' RANKS AND SALARIES

School and rank	*First year*	*Second year*	*Third year*	*Fourth year*	*Fifth year*	*Sixth year*	*Seventh year*	*Eighth year*	*Ninth year*
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Supervision									
SUPERVISORS									
Drawing	2150	2300	2400	2500	2700	2850	3000	....	....
Woman	....	....	....	....	....	....	**1**	....	....
Kindergarten	2150	2300	2400	2500	2700	2850	3000	....	....
Woman	....	....	....	....	....	....	**1**	....	....
Manual Training	2150	2300	2400	2500	2700	2850	3000	....	....
Man	....	....	....	....	**1**	....	....	....	....
Music	2150	2300	2400	2500	2700	2850	3000	....	....
Man	....	....	....	....	**1**	....	....	....	....
Penmanship	2120	2300	2400	2500	2700	2850	3000	....	....
Man	....	....	....	....	**1**	....	....	....	....
Physical Training	2150	2300	2400	2500	2700	2850	3000	....	....
Man	....	....	....	....	....	**1**	....	....	....
Primary	1400	1500	1640	1720	1800	1900	2000	....	....
Women	....	....	....	....	....	....	**3**	....	....
School Gardens (12 months)	1800	....	....	....	....	....	....	....	....
Man	**1**	....	....	....	....	....	....	....	....
Special Schools	1640	....	....	....	....	....	....	....	....
Woman	**1**	....	....	....	....	...	....	....	....
HIGH SCHOOLS									
Assistant Principal	2150	2300	2400	2500	2700	2850	3000	....	....
Men	....	**1**	....	....	....	....	**3**	....	....
Head Assistant	2000	2040	2100	2180	....	....	....	....	....
Men	**4**	**7**	**1**	**26**	....	....	....	....	....
Women	....	**6**	....	....	....	....	....	....	....
First Assistant	1640	1700	1800	1900	2000	....	....	....	....
Men	**4**	**15**	**5**	**3**	**19**	....	....	....	....
Women	**7**	**10**	**2**	**5**	**3**	....	....	....	....
Second Assistant	1120	1180	1240	1300	1360	1460	1520	1580	1640
Men	**1**	**3**	**5**	**4**	**8**	**22**	**1**	**3**	**18**
Women	**7**	**32**	**3**	**6**	**8**	**13**	**2**	**7**	**11**
Substitute Assistant	980	1020	....	....	....	....	....	....	....
Men	**2**	**1**	....	....	....	....	....	....	....
Women	**9**	**10**	....	....	....	....	....	....	....
Substitute (2 A Gr.)	1072	....	....	....	....	....	....	....	....
Woman	**1**	....	....	....	....	....	...	....	....
Instructor Phys. Tr. (Boys)	1200	1300	1400	1500	1600	1700	1800	....	....
Men	....	....	....	**2**	....	**1**	**3**	....	....
Instructor Phys. Tr. (Girls)	1100	1200	1300	1400	1500	....	....	....	....
Women	**1**	**2**	....	**3**	**4**	....	....	....	....
Clerk 2 A Gr. + $100 (for 12 mos.)	700	740	800	900	1000	1072	1132	....	....
Women	....	....	**1**	....	....	**2**	**5**	....	....
Clerk, Summer, 2 A Gr.	600	640	700	800	900	972	1032	....	....
Woman	....	**1**	....	....	....	....	....	....	....
ELEMENTARY SCHOOLS									....
Head Assistant	1180	1240	1300	....	....	....	....	....	....
Women	**16**	**1**	**47**	....	....	....	....	....	....
First Assistant	920	1020	1072	1120	....	....	....	....	....
Women	....	....	**123**	**173**	..	....	....	....	....
Second Assistant	600	640	700	800	900	972	1032	....	....
Women	**11**	**88**	**77**	**75**	**79**	**72**	**616**	....	....
(Substitutes) Permanent	**110**	**17**	**3**	**1**	....	....	....	....	....
Temporary	**8**	**2**	....	....	....	....	....	....	....

pactly as is consistent with clearness. Table 56, from the 1914–15 *Report of Board of Education* in St. Louis, does this very suggestively by indicating in one table the number of years in the salary schedule for each position, the corresponding salary for each year and for each grade, and the number of men and number of women who draw the salaries stated for each grade.

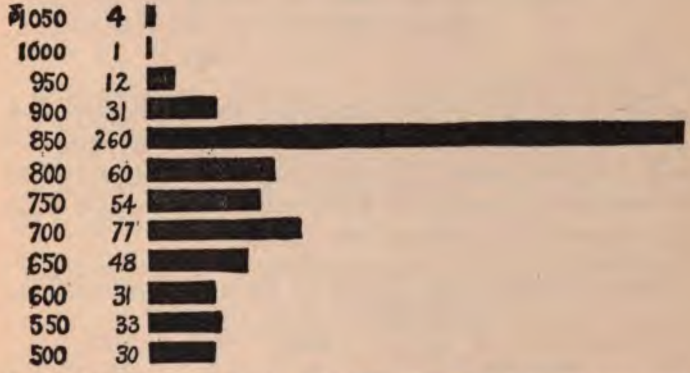


Diagram 58. Showing Number of Elementary Teachers receiving Various Salaries

Showing the salary situation. To picture the general salary situation clearly the Rochester *School Report* (1911–13), p. 66, makes use of a bar diagram to good effect, as reproduced in Diagram 58.

The growth of salaries in the system, as shown by average salaries paid and by the corresponding percentage of increase for each grade of position during past years, may similarly be shown in a table.

The general level of teaching salaries may also be brought out by some such tabular representation as that shown in Table 62.

TABLE 57. SHOWING THE GENERAL LEVEL OF SALARIES IN A CITY

Salary	*Teachers*
$3000 or over	1
2000 to $2999	4
1500 to 1999	24
1000 to 1499	30
800 to 999	30
600 to 799	238
400 to 599	8
200 to 399	8
	343

The *years of teaching experience* should be reported, *first*, for *total experience*, and, *second*, for years of *experience within the local system*. Table 58 suggests a compact tabular arrangement for such data, which can be used to show either type of statistical information.

TABLE 58. SHOWING THE YEARS OF TEACHING SERVICE OF ALL TEACHERS EMPLOYED IN 1915–16

	Grades	*H. S.*	*Total*
Beginners	11	6	17
2 years	2	..	2
3 to 4 years	4	1	5
5 to 9	2	1	3
10 to 14	2	..	2
15 to 19	2	..	2
20 to 24	2	1	3
25 to 29	1	..	1
Total	26	9	35

The training of the teachers. This should be reported by the same sort of tabular arrangement as that showing the salary distribution. Diagram 59 suggests a graphic method[1] by which this, as well as many other kinds of school facts may be reported.

---

[1] Jessup, W. A. *The Teaching Staff*, p. 58. (Cleveland Education Survey Monographs.)

DIAGRAM 59. PER CENT OF ELEMENTARY TEACHERS, HIGH-SCHOOL TEACHERS, AND ELEMENTARY PRINCIPALS IN CLEVELAND WHO ARE HOME-TRAINED AND NOT HOME-TRAINED

(After Jessup, 1916.)

Size of classes. The size of classes within the local system should be reported annually. Table 59 shows in simple form the size of classes in the school system as a whole, while Table 60 shows the size of classes in each main division of the school system.

TABLE 59. SHOWING THE NUMBER OF PUPILS PER TEACHER, ELEMENTARY GRADES, DECEMBER, 1910

23 teachers had	over 50 pupils
90	45 to 49
63	40 to 44
56	35 to 39
14	30 to 34
8	below 30

TABLE 60. SHOWING THE NUMBER OF PUPILS PER TEACHER IN DIFFERENT CLASSES OF SCHOOLS

	Auxiliary school	*High school*	*Grammar grades*	*Primary grades*	*Kindergarten*
1909–10....	68	22.2	34.0	36.0	32.4
1910–11....	70	19.7	32.2	35.8	30.3
1911–12....	96	19.8	31.5	35.9	32.3
1912–13....	93	19.2	27.5	31.7	27.6
1913–14....	150	19.5	27.2	32.3	27.2
1914–15....	..	23.4	33.9	35.3	35.0

Often it is desirable to show the size of classes in the city, compared with those in other city school systems of the same size or class. In such cases Table 61 gives a good form of table for displaying such information.

Table 61. Number of Pupils in Average Daily Attendance per Teacher in Elementary Schools in 19 American Cities, 1915

(Data from Annual Report, United States Commissioner of Education, 1915, vol. 2)

	No. of teachers employed	*Average daily attendance*	*No. of pupils per teacher*	*Rank*
Albany	320	9,427	29.5	6
Birmingham	571	17,781	31.1	8
Bridgeport	388	15,093	38.9	18
Cambridge	386	12,255	32.0	10
Dayton	433	13,242	30.6	7
Des Moines	486	13,021	27.0	2
Fall River	499	12,899	25.8	1
Grand Rapids	471	12,909	*Prim.* 32.3 *Gram.* 27.2	10 4
Kansas City	337	11,026	32.7	12
Lowell	264	9,665	36.6	15
Lynn	285	10,793	37.8	17
Memphis	450	14,070	31.3	9
Nashville	314	14,135	44.7	19
New Bedford	353	11,466	32.5	11
Paterson	462	17,362	37.6	16
Richmond	599	20,142	33.6	14
San Antonio	373	10,253	27.5	5
Scranton	550	18,014	32.8	13
Springfield	490	13,296	27.2	3

4. The reporting of facts concerning the pupil

Data needed, and forms. There are eight types of fact that the annual school report should give the public and school officers about the pupil: (1) the number of children of school census age in the city; (2) the total enrollment in all schools in the city; (3) the total enrollment in public schools; (4) as closely as possible, the estimated enrollment in parochial schools; (5) the total and the average enrollment and average daily attendance in each of the various grades, kindergarten to last grade in high school inclusive; (6) the distribution of children in each grade according to age; (7) the distribution of children in each grade according to number of years spent in the grade; (8) distribution of children in each grade according to the facts concerning "promotion." Tables 62 to 67 suggest tabular arrangements of these data.[1]

Picturing the holding power of the schools. It will be desirable to use graphic devices to picture the efficiency with which the school machinery holds pupils in school, grades and classifies them, and promotes them through the various grades. Diagram 60 represents a good method of presenting to the people the degree to which the public schools are educating the children of school age in the city.

Diagram 61 is an excellent pictorial device, taken from the 1914–15 St. Louis *School Report*, for showing the increase in persistence of children in school. Such a diagram is clear and is easily comprehended by citizens. Diagram 62 [2] suggests a method of illustrating the "holding power" of the schools. Diagrams 5, 6, and 7, in Chapter I, give graphic methods of studying and reporting failures in the schools, by grades and by subjects.

5. Reporting facts as to the school plant

School buildings. The following topical list of points should be covered in reporting facts as to the school buildings in use: —

1. Number of buildings, — elementary, intermediate, secondary, covering a period of years.
2. Number of classrooms in use at stated time, covering comparison of several years.
3. Valuation of school property; historical, several years.

---

[1] The writer is indebted to Dr. L. P. Ayres for the material in Tables 62 to 67 inclusive.

[2] Ayres, L. P., *Child Accounting in the Public Schools*, p. 19. (Cleveland Education Survey Monographs.)

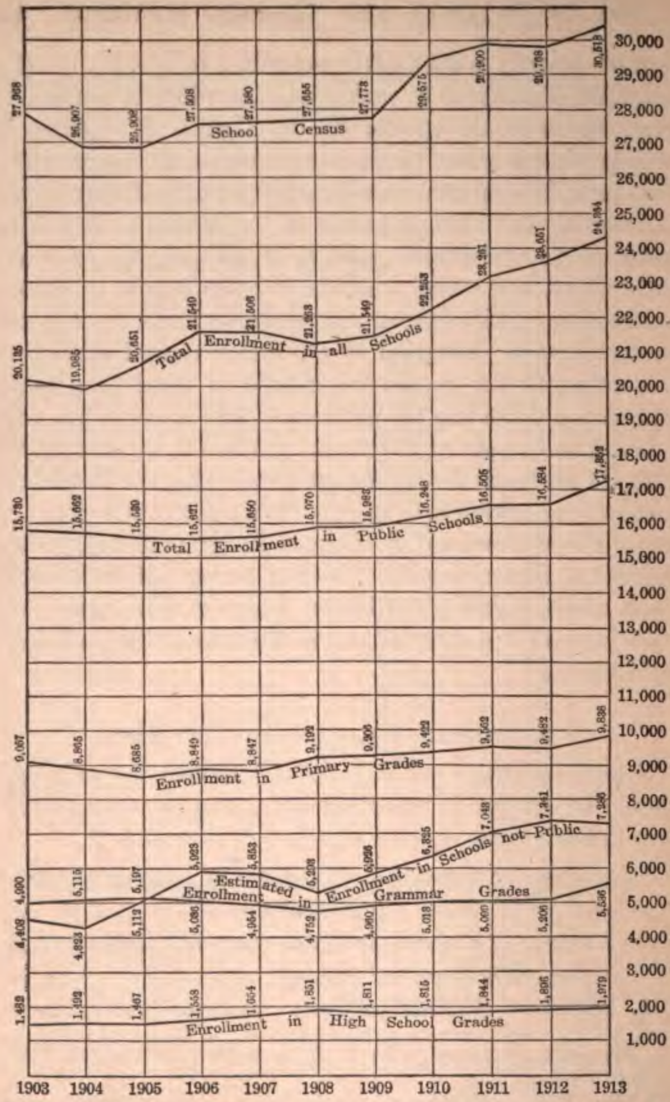


DIAGRAM 60. SHOWING FOR A SERIES OF YEARS THE DEGREE TO WHICH PUBLIC SCHOOLS ARE EDUCATING THE CHILDREN OF SCHOOL AGE IN THE CITY

(*Grand Rapids School Report, 1915.*)

4. Cost of new buildings. A tabular form for such data is suggested by Table 68.
5. Standards used in judging buildings. For such facts a graphic form is shown in Diagrams 69 and 70.

Standards may be set, as shown in Diagram 70, against the individual buildings of the system to permit of a judgment as to the present condition of the school plant.

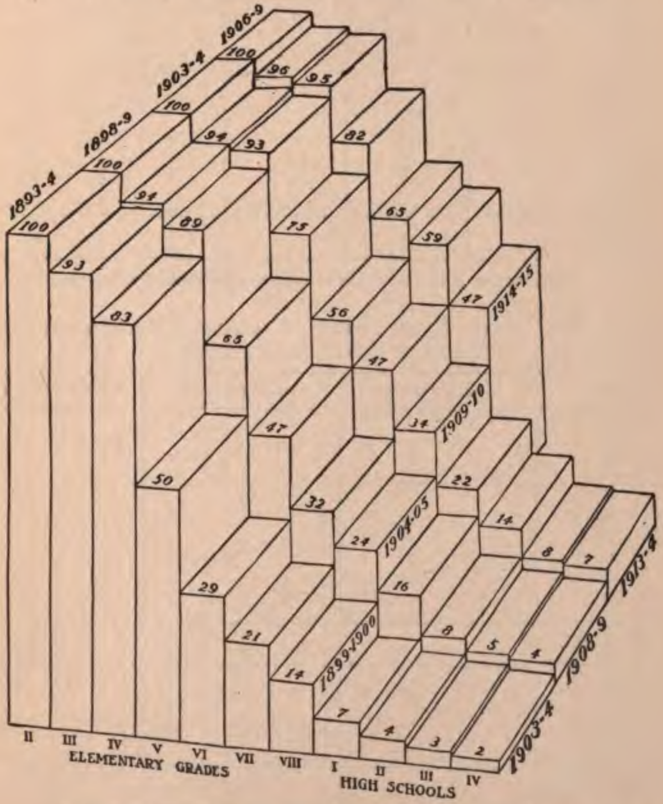


DIAGRAM 61. PERSISTENCE OF ATTENDANCE AT SCHOOL

(From the St. Louis *School Report*, 1914–15.) Numbers on top of columns show the number of pupils out of each 100 entering the second grade in the years indicated who were enrolled in the several grades in the succeeding years.

TABLE 62. NUMBER OF CHILDREN OF SCHOOL CENSUS AGE — ILLUSTRATIVE FORM

Age	*Public schools*	*Private schools*	*Parochial schools*	*In no schools*	*Total*
5					
6					
......	....	....	....	....	....
7					
8					
9					
10					
11					
12					
13					
......	....	....	....	....	....
14					
15					
16					
17					
Total *					

* Data between dotted lines refer to children of compulsory school age.

TABLE 63. TOTAL ENROLLMENT, AVERAGE ENROLLMENT, AND AVERAGE ATTENDANCE, 1916–17

Grade	*Total enrollment*	*Average enrollment*	*Average attendance*
K			
1			
2			
3			
4			
5			
6			
7			
8			
Total El.			
I			
II			
III			
IV			
Total High			
Night			

TABLE 64. AGES OF CHILDREN IN EACH GRADE AS OF JUNE, 1917

Grade	*Ages*												*Total*	*Over age*	*Per cent over age*
	6	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*			
1	117	187	49	9	3	4							369	65	18
2	2	153	201	47	23	5	1						432	76	18
3		2	116	219	86	27	10	2	1				463	126	27
4			18	112	169	78	37	22	7	1	1		445	146	33
5				6	80	150	85	57	27	11	4		420	184	44
6					6	80	114	85	59	24	6		374	174	47
7						6	66	114	107	65	12	1	371	185	50
8							4	47	107	96	30	5	289	131	45
Total	119	342	384	393	367	350	317	327	308	197	53	6	3163	1087	34

TABLE 65. SHOWING YEARS IN SCHOOL OF CHILDREN IN EACH GRADE, AS OF JUNE, 1917

Grade	*Years in school*											*Total*	*Slow*	*Per cent slow*
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*			
1	322	37	7	2	1							369	47	13
2	65	261	82	21	2	1						432	106	25
3		67	266	95	27	6	1	1				463	130	28
4		3	81	226	91	30	12	2				445	135	30
5				53	177	109	53	22	3	1	2	420	190	45
6				5	67	157	76	52	14	3		374	145	39
7					3	55	178	91	34	6	4	371	135	36
8					2	4	23	164	76	20		289	96	33
Total	387	368	436	402	370	362	343	332	127	30	6	3163	984	31

TABLE 66. SHOWING ATTENDANCE IN ELEMENTARY SCHOOLS DURING 1916–17

Days attended	*Pupils*	*Days attended*	*Pupils*
0– 9		100–109	
10–19		110–119	
20–29		120–129	
30–39		130–139	
40–49		140–149	
50–59		150–159	
60–69		160–169	
70–79		170–179	
80–89		180–189	
90–99		190–199	
		200	
Total		Total	

TABLE 67. SHOWING PROMOTIONS FOR SCHOOL YEAR ENDING JUNE, 1917

Grade	*On June promotion list*	*Uncon-ditionally promoted*	*Con-ditionally promoted*	*Left behind*	*Promoted more than one grade*	*Special promotion between September and June*	*Number who were promoted and dropped back*
K							
1							
2							
3							
4							
5							
6							
7							
8							
Total							
I							
II							
III							
IV							
Total							

TABLE 68. COST DATA FOR NINE FIREPROOF ELEMENTARY-SCHOOL BUILDINGS IN BOSTON

School	*Date*	*Class-rooms*	*Special rooms*	*Pupils*	*Cost of building*	*Plans, specifications, and inspection*	*Cost per classroom*	*Cost per pupil*	*Cost per cubic foot*
Patrick Collins	1907	17	5	753	$194,330	$17,666	$11,431	$258	$.268
Edward Everett	1909	14	7	512	118,267	10,752	8,448	231	.229
Nathan Hale	1909	12	4	400	73,955	6,723	6,163	185	.222
John Cheverus	1909	16	7	587	112,977	10,271	7,061	192	.211
Peter Faneuil	1910	17	4	633	118,888	10,808	6,993	188	.275
William Lloyd Garrison	1910	10	2	383	72,766	6,615	7,277	190	.264
Samuel Adams	1910	22	7	757	158,194	14,381	7,191	209	.260
Lafayette	1911	8	4	293	69,084	6,280	8,636	236	.313
Abraham Lincoln	1911	40	8	1,517	308,097	28,009	7,702	203	.266

6. *Reporting miscellaneous educational information*

The foregoing sections have presented, in outline, definite suggestions for the content and form of the school report on five principal phases: (1) the legal basis; (2) school finance; (3) the teaching staff; (4) the pupil; (5) school buildings. The primary aim has been to illustrate, in compact form, suggestive tabular and graphic means for setting forth effectively such information.[1] At the same time the actual facts needed in the report have been sketched. In addition, there are many other types of miscellaneous school facts that

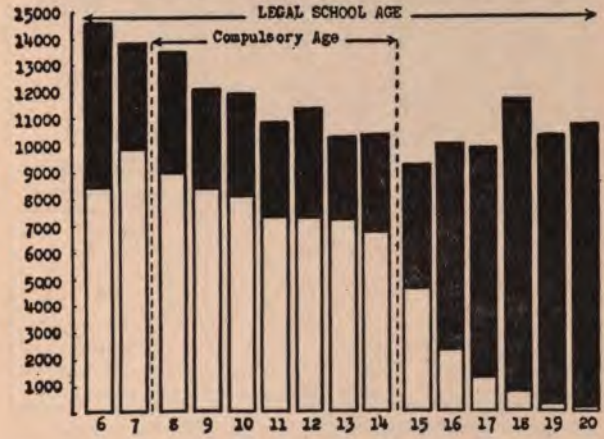


DIAGRAM 62. SHOWING THE HOLDING POWER OF THE SCHOOLS

The columns represent the children enumerated by the school census as of each age from 6 through 20. Portion in outline represents children in public schools. Portion in black represents those not in public schools. (*Cleveland Education Survey Report*, 1916.)

---

[1] A very complete compilation of *Graphic Methods for Presenting Facts* has been published by W. C. Brinton in a book by that name. (Engineering Magazine Company, New York, 1914.) The preparer of graphic reports, in whatever subjects, will receive very great aid from consulting this book.

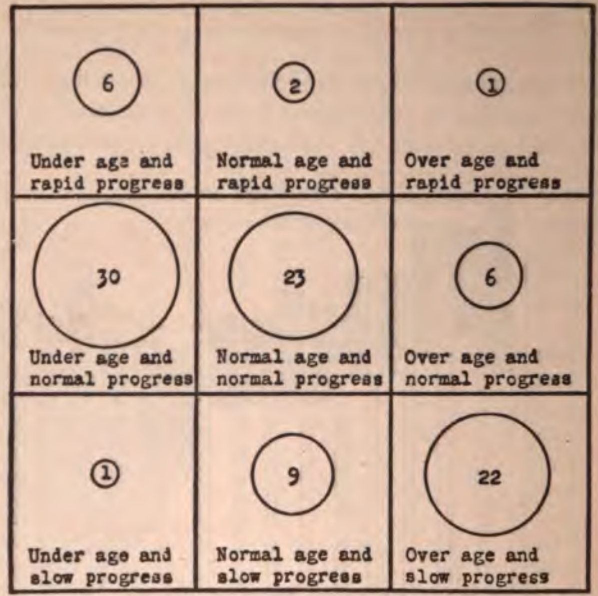


DIAGRAM 63. PER CENT OF CHILDREN IN EACH AGE AND PROGRESS GROUP IN ELEMENTARY SCHOOLS AT CLOSE OF YEAR 1914–15

(*Cleveland Education Survey Report*, 1916.)

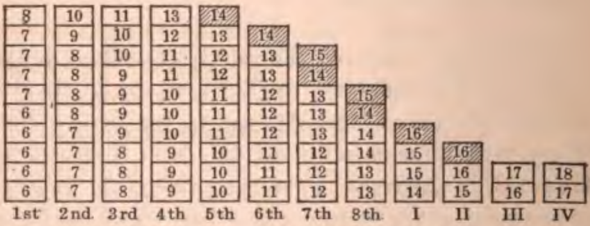


DIAGRAM 64. PROGRESS OF TEN TYPICAL PUPILS THROUGH THE SCHOOL SYSTEM

Each square represents one child. The number represents his age. As they advance through the grades, they advance in age. The shaded squares represent those who drop out. (*Cleveland Education Survey Report*, 1916.)

should be tabulated and graphed, the full presentation of which must be left to a volume devoted to the specific problem of this chapter. It may be of some service to school men,

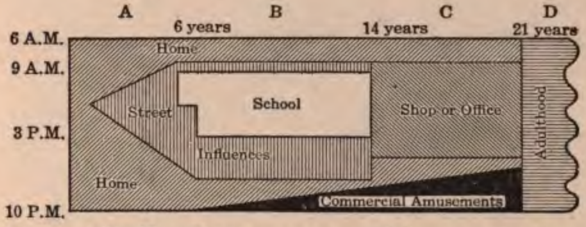


DIAGRAM 65. THE ENVIRONMENT OF A MINOR DURING THE PRINCIPAL PERIODS OF HIS GROWTH

(From Perry, C. A., *Educational Extension*, p. 35.)

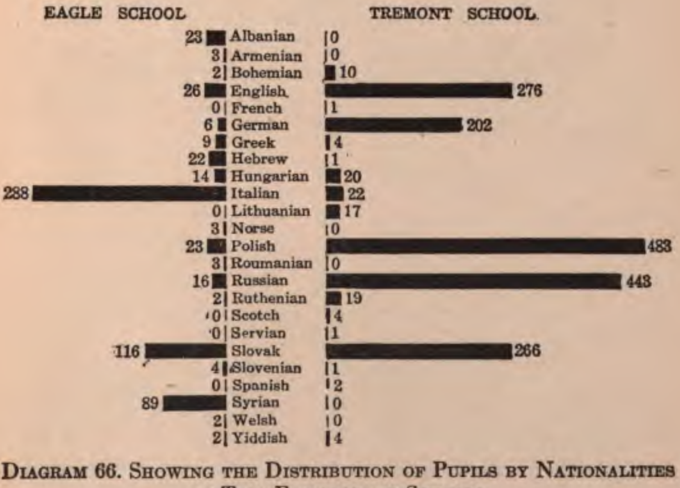
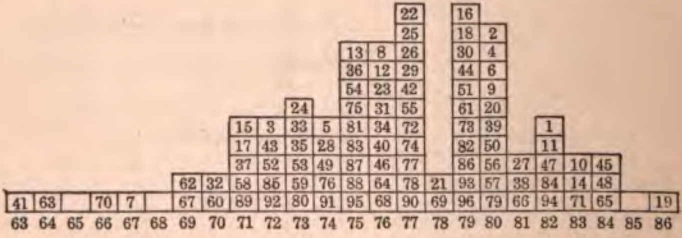


DIAGRAM 66. SHOWING THE DISTRIBUTION OF PUPILS BY NATIONALITIES IN TWO ELEMENTARY SCHOOLS

(*Cleveland Education Survey Report*, 1916.)

DIAGRAM 67. HOW 915 CHILDREN SPENT THEIR SPARE TIME ON TWO PLEASANT DAYS IN JUNE

(Johnson, G. E., *Education through Recreation*, p. 47. *Cleveland Education Survey Report*, 1916.)



														22		16							
														25		18	2						
												13	8	26		30	4						
												36	12	29		44	6						
												54	23	42		51	9						
										24		75	31	55		61	20						
								15	3	33	5	81	34	72		73	39		1				
								17	43	35	28	83	40	74		82	50		11				
								37	52	53	49	87	46	77		86	56	27	47	10	45		
						62	32	58	85	59	76	88	64	78	21	93	57	38	84	14	48		
41	63		70	7		67	60	89	92	80	91	95	68	90	69	96	79	66	94	71	65		19
63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86

DIAGRAM 68. AVERAGE SCORES MADE IN SPELLING BY THE NINETY-SIX ELEMENTARY SCHOOLS

The figures below the diagram show the percentages, and the ones in the diagram show the numbers of the schools. (From Judd, C. H., *Measuring the Work of the Public Schools*, p. 84. *Cleveland Education Survey Report*, 1916.)

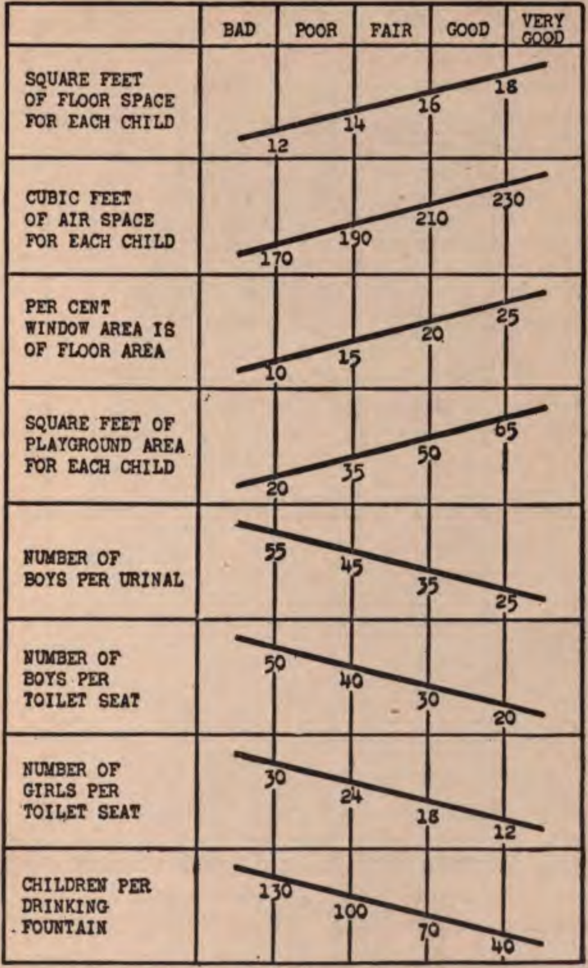
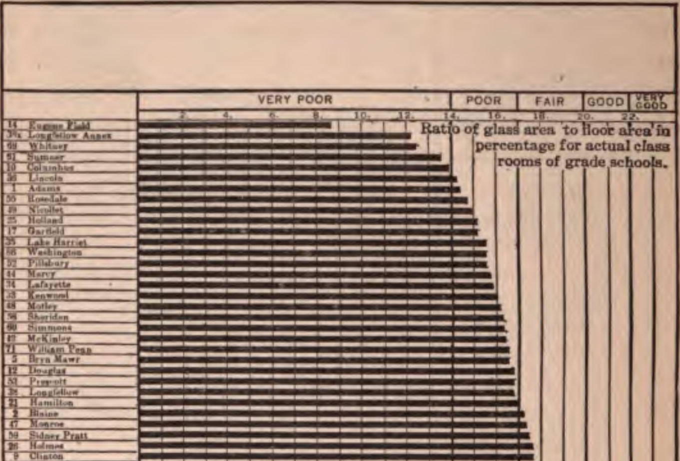


DIAGRAM 69. SOME STANDARDS USED IN JUDGING SCHOOL BUILDINGS

(Ayres, L. P., *School Buildings and Equipment*, p. 54. *Cleveland Education Survey Report*, 1916.)



Note;- The Minnesota Department of Education requires a minimum glass area equal to 20% of the floor area for elementary class rooms.

DIAGRAM 70. RATIO OF GLASS AREA TO FLOOR AREA

(From *A Million a Year.* Minneapolis Board of Education, 1916.)

however, in closing this volume, to bring together in Diagrams 71 to 80, inclusive, a few striking pictorial methods of presenting such miscellaneous school facts which have been used effectively by school men, in presenting information to the people of their school city. Diagrams 74 to 78 inclusive, and Diagram 80, have been quoted from *Help-Your-Own-School Suggestions*, Bulletin No. 31, Feb. 21, 1914, of the Bureau of Municipal Research, New York City. The others

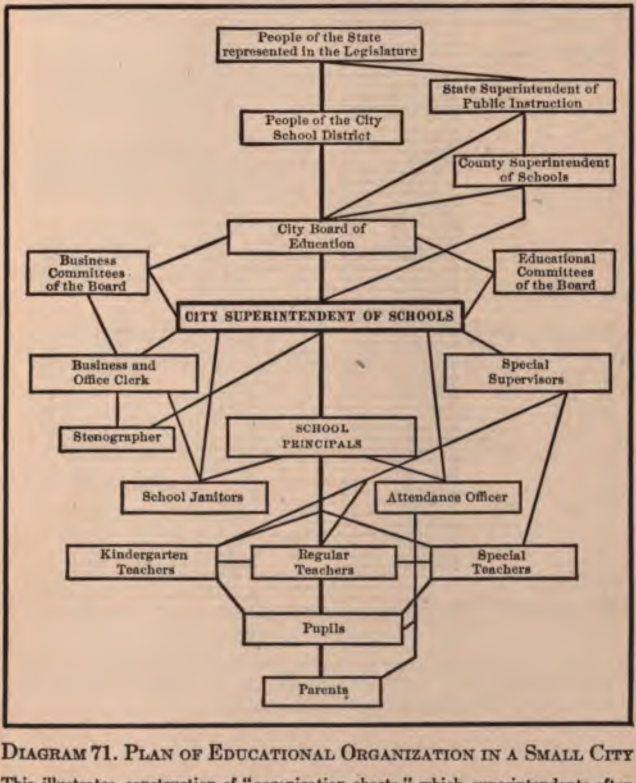


DIAGRAM 71. PLAN OF EDUCATIONAL ORGANIZATION IN A SMALL CITY

This illustrates construction of "organization charts," which superintendents often desire to show. (From Cubberley, E. P., *Public School Administration*, p. 167.)

are properly credited to the report from which they have been taken.

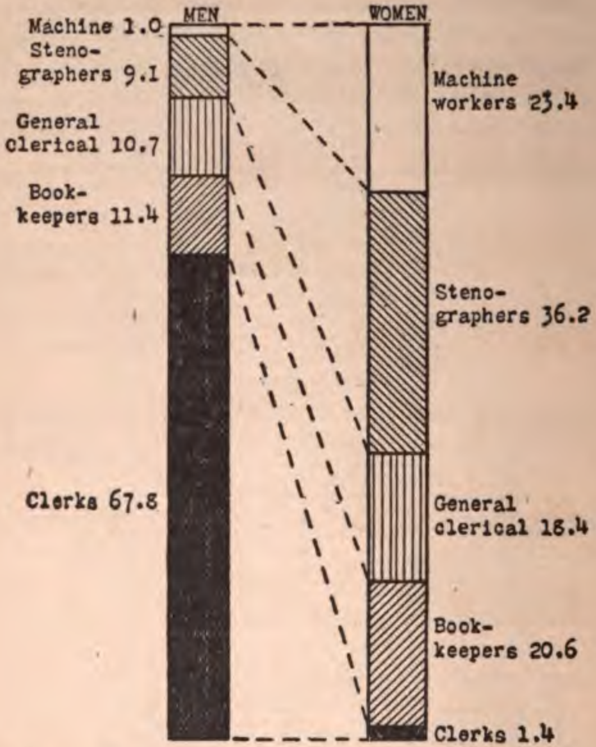


Diagram 72. Percentage Distribution of Non-Administrative Positions in Office Work

As held by men and women in Cleveland, 1912–15, 1955 positions for men and 2747 for women. (From Stevens, B. E., *Boys and Girls in Commercial Work*, p. 26. *Cleveland Education Survey Report*, 1916.)

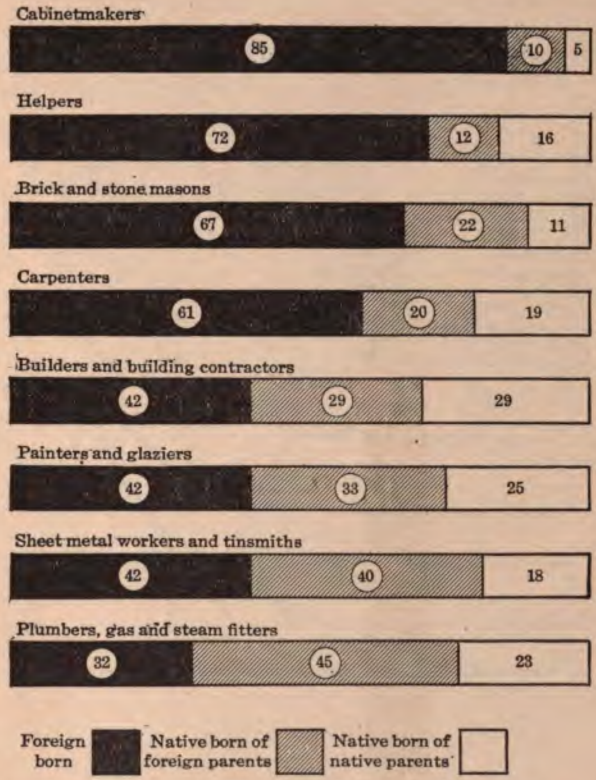


DIAGRAM 73. PERCENTAGE OF WORKERS IN BUILDING TRADES THAT ARE FOREIGN-BORN, NATIVE-BORN OF FOREIGN PARENTS, AND NATIVE-BORN OF NATIVE PARENTS

(From Shaw, F. L., *The Building Trades*, p. 33. *Cleveland Education Survey Report*, 1916.)

SPELLING MISFITS

DIAGRAM 74. ILLUSTRATING SPELLING DIFFICULTIES

DIAGRAM 75. ILLUSTRATING THE IMPORTANCE OF AFTER-SCHOOL ACTIVITIES

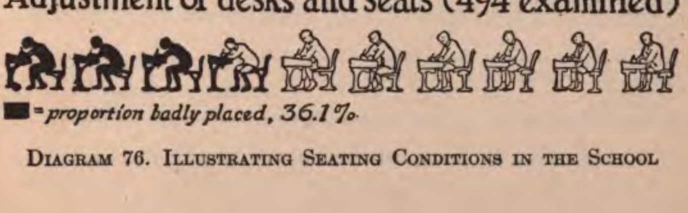


DIAGRAM 76. ILLUSTRATING SEATING CONDITIONS IN THE SCHOOL

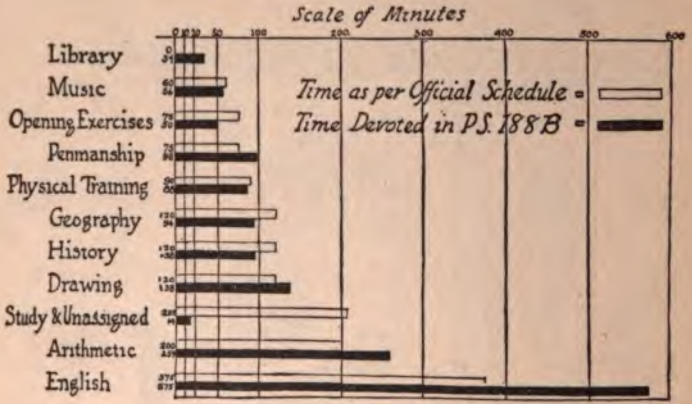


DIAGRAM 77. ILLUSTRATING THE SCHOOL PROGRAM

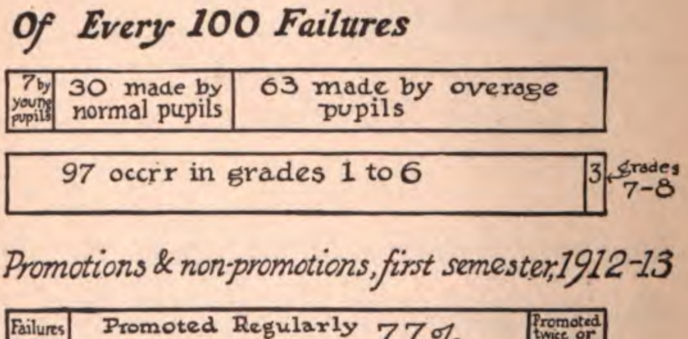


DIAGRAM 78. ILLUSTRATING PROMOTION AND FAILURES

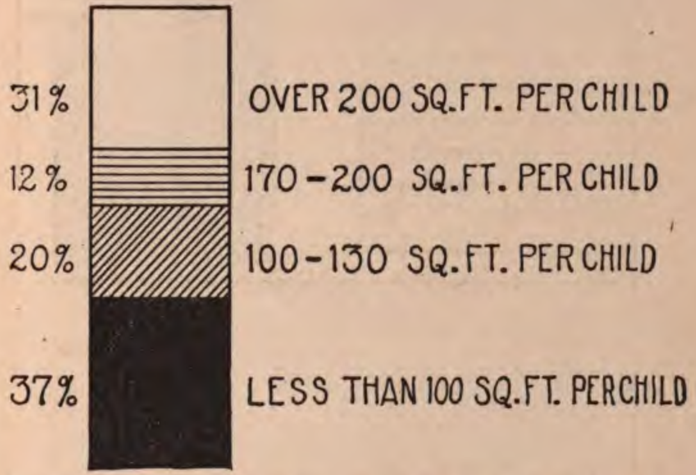


DIAGRAM. 79 SHOWING THE PERCENTAGE OF CHILDREN HAVING PLAYGROUNDS OF VARIOUS SIZES

(*Salt Lake City School Survey Report*, 1915.)

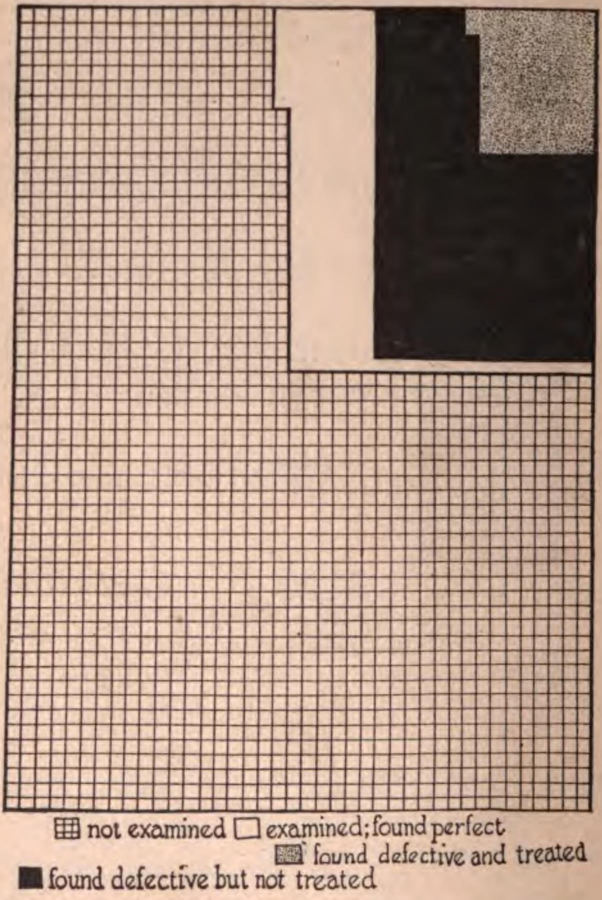


Diagram 80. What the School Records relating to Medical Examinations show

BIBLIOGRAPHY OF THE QUANTITATIVE STUDIES ON SCHOOL ADMINISTRATION

For Tabular Key see Plate I.

I. STUDIES ON THE COURSE OF STUDY

1. Ayres, L. P. *A Scale for the Measurement of Spelling Ability.* Russell Sage Foundation, Division of Education, 1915.
2. Bagley, W. C., and Rugg, H. O. *The Content of American History as Taught in the Seventh and Eighth Grades.* University of Illinois, School of Education, Bulletin No. 16, 1916.
3. Baker, J. H., Chairman. *Economy of Time in Education.* 106 pp. United States Bureau of Education, Bulletin No. 38, 1913.
4. Bobbitt, J. F., and others. "Literature in the Elementary Curriculum." *Elementary School Teacher,* December, 1913, and January, 1914.
5. Cleveland Education Survey Reports, 1916. These reports can be secured from the Survey Committee of the Cleveland Foundation, Cleveland, Ohio, and from the Division of Education of the Russell Sage Foundation, New York City.
    1. *Child Accounting in the Public Schools* — Ayres.
    2. *Educational Extension* — Perry.
    3. *Education through Recreation* — Johnson.
    4. *Financing the Public Schools* — Clark.
    5. *Health Work in the Public Schools* — Ayres.
    6. *Household Arts and School Lunches* — Boughton.
    7. *Measuring the Work of the Public Schools* — Judd.
    8. *Overcrowded Schools and the Platoon Plan* — Hartwell.
    9. *School Buildings and Equipment* — Ayres.
    10. *Schools and Classes for Exceptional Children* — Mitchell.
    11. *School Organization and Administration* — Ayres.
    12. *The Public Library and the Public Schools* — Ayres and McKinnie.
    13. *The School and the Immigrant* — Miller.
    14. *The Teaching Staff* — Jessup.
    15. *What the Schools Teach and Might Teach* — Bobbitt.
    16. *The Cleveland School Survey* (summary) — Ayres.
    17. *Boys and Girls in Commercial Work* — Stevens.
    18. *Department Store Occupations* — O'Leary.
    19. *Dressmaking and Millinery* — Bryner.
    20. *Railroad and Street Transportation* — Fleming.
    21. *The Building Trades* — Shaw.
    22. *The Garment Trades* — Bryner.
    23. *The Metal Trades* — Lutz.
    24. *The Printing Trades* — Shaw.
    25. *Wage-Earning and Education* (summary) — Lutz.

6. Cook, W. A., and O'Shea, M. J. *The Child and His Spelling.* Bobbs Merrill Co., Indianapolis, Indiana.
7. Iowa State Teachers' Association, Bulletin of the, November, 1916. *Elimination of Obsolete and Useless Topics and Materials from the Common Branches.* G. W. Wilson, Chairman.
8. Jessup, W. A., and Coffman, L. D. *The Supervision of Arithmetic.*
9. Jones, W. F. *Concrete Investigation of the Material of English Spelling.* Published by the University of South Dakota, Vermilion, South Dakota, 10 cents.
10. Koos, L. V. *The Administration of Secondary School Units.* University of Chicago Press, 1917. No. 3 of the Supplementary Educational Monographs.
11. Minnesota Education Association, March, 1914. *Report of the Committee on the Elementary Course of Study.*
12. National Education Association. *1916 Report of the National Council of Education Committee on Standards and Tests of Efficiency.* Fifteenth Yearbook of National Society for the Study of Education, 1916. Part I.
13. The Fourteenth Yearbook of the National Society for the Study of Education, 1915. Part I: *Minimal Essentials in the Elementary School Subjects — Standards and Current Practices.*
14. The Sixteenth Yearbook of the National Society for the Study of Education, 1917. Part I: *Second Report of the Committee on Minimal Essentials in the Elementary School Subjects.*
15. The Fifteenth Yearbook of the National Society for the Study of Education, 1916. Part III: *The Junior High School,* A. A. Douglas.
16. Payne, B. R. *Elementary School Curricula.* Silver, Burdett & Co., 1895.
17. Rice, J. M. *Scientific Management in Education.* Hinds, Noble & Eldridge, New York, 1913.

17*a*. Rugg, H. O., and Clark, J. R. "Standardized Tests and the Improvement of Teaching in First-Year Algebra." *School Review,* February, March, May, and October, 1917.

18. Studley, C. K., and Ware, Allison. *Common Essentials in Spelling. Chico State Normal School, Chico, California,* Bulletin No. 7, 1914.

II. QUANTITATIVE STUDIES OF THE TEACHING STAFF AND TEACHING EFFICIENCY

(Studies dealing with the social status, training, experience, tenure, selection, and salaries of teachers.)

19. Ballou, F. W. *Appointment of Teachers in Cities.* Harvard Studies in Education, vol. II, 1915. Harvard University Press.

20. Boice, A. C. *Methods of Rating Teaching Efficiency.* Part II, Fourteenth Yearbook of the National Society for the Study of Education, 1915. (See also "Qualities of Merit in Secondary School Teachers," *Journal of Educational Psychology,* March, 1912.)

21. Coffman, L. D. *The Social Composition of the Teaching Population.* Teachers College, Columbia University, Contributions to Education, 1911, No. 41.

22. Cubberley, E. P. *The Certification of Teachers.* Part II, Fifth Yearbook of National Society for the Study of Education. University of Chicago Press, 1906.

23. Elliott, E. C. *A Provisional Plan for the Measure of Merit of Teachers.* Department of Public Instruction, Madison, Wisconsin, 1910.

24. Hood, W. R. *Digest of State Laws Relating to Public Education.* United States Bureau of Education, Bulletin No. 47, 1915.

25. Judd, C. H., and Parker, S. C. *Problems Involved in Standardizing State Normal Schools.* United States Bureau of Education, Bulletin No. 12, 1916.

26. Littler, S. "Causes of Failure Among Elementary School Teachers." *School and Home Education,* March, 1914.

27. Manny, F. A. *City Training for Teachers.* United States Bureau of Education, Bulletin No. 47, 1914.

28. Moses. "Causes of Failure Among High-School Teachers." *School and Home Education,* January, 1914.

29. National Education Association. *Report of the Committee on Salaries, Tenure and Pensions,* 1905.

30. National Education Association. *Report of the Committee on Teachers' Salaries and Cost of Living,* January, 1913.

31. Ruediger, W. C. *Agencies for the Improvement of Teachers in Service.* United States Bureau of Education, Bulletin No. 3, 1911.

32. Ruediger, W. C., and Strayer, G. D. "The Qualities of Merit in Teachers." *Journal of Educational Psychology*, May, 1910, vol. I, pp. 272–78.
33. *Salaries of Teachers and School Officers, A Comparative Study of*. United States Bureau of Education, Bulletin No. 31, 1915.
34. *The Tangible Rewards of Teaching*. United States Bureau of Education, Bulletin No. 16, 1914. Compiled by J. C. Boykin and Roberta King for the Committee of the National Education Association on Teachers, Salaries, and Cost of Living.
35. Thorndike, E. L. *The Teaching Staff of Secondary Schools in the United States*. United States Bureau of Education, Bulletin No. 4, 1909.
36. Updegraff, H. *Teachers' Certificates Issued Under General State Laws and Regulations*. United States Bureau of Education, Bulletin No. 18, 1911.

III. QUANTITATIVE STUDIES DEALING WITH THE PUPIL

A. ELIMINATION AND RETARDATION

37. Ayres, L. P. *Laggards in Our Schools*. Russell Sage Foundation, Division of Education, New York City, 1909.
38. Bachman, F. P. *Problems in Elementary School Administration*. World Book Company, Yonkers, New York, 1914.
39. Blan, A. B. *A Special Study of the Incidence of Retardation*. Teachers College, Columbia University, Contributions to Education, No. 40, 1911.
40. Dearborn, W. F. "The Qualitative Elimination of Pupils from School." *Elementary School Teacher*, September, 1910.
41. Holley, C. E. *The Relationship between Persistence in School and Home Conditions*. Part II, Fifteenth Yearbook, National Society for the Study of Education, 1916.
42. Keyes, C. H. *Progress through the Grades of City Schools*. Teachers College, Columbia University, Contributions to Education, No. 42, 1911.
43. Strayer, G. D. *Age and Grade Census of Schools and Colleges: A Study of Retardation and Elimination*. United States Bureau of Education, Bulletin No. 5, 1911.
44. Thorndike, E. L. *Elimination of Pupils from School*. United States Bureau of Education, Bulletin No. 4, 1907.

45. Van Den Burg, J. K. *Causes of Elimination of Students in Public Secondary Schools of New York City.* Teachers College, Columbia University, Contributions to Education, No. 47, 1911.

B. SIZE OF CLASS AND EFFICIENCY OF INSTRUCTION

46. Breed, F. S., and McCarthy, J. D. "Size of Class and Efficiency of Teaching." *School and Society*, vol. IV, No. 104, December 23, 1916, pp. 965–71.
47. Boyer, P. A. "Class Size and School Progress." *Psychological Clinic*, vol. VIII, 1914.
48. Cornman, O. P. "Size of Classes and School Progress." *Psychological Clinic*, vol. III, 1909.
49. Harlan, C. L. "Size of Class as a Factor in School-Room Efficiency." *Journal of Educational Administration and Supervision*, vol. I, 1915.

C. GRADING, CLASSIFICATION AND PROMOTION OF PUPILS

50. Bunker, F. F. *Reorganization of the Public School System.* United States Bureau of Education, Bulletin No. 8, 1916.
51. Burk, F. Monograph A, *A Remedy for Lock-Step Schooling.* 15 cents 1913. Monograph C, *Every Child* vs. *Lockstep Schooling.* 15 cents. 1915. San Francisco State Normal School, California.
52. Quantitative and descriptive historical summaries on "promotion plans" in *Annual Reports* of the United States Commissioner of Education: (1) 1890–91, p. 991; (2) 1891–92, pp. 600–32; (3) 1898–99, p. 330.
53. Dearborn, W. F. *The Relative Standing of Pupils in the High School and University.* University of Wisconsin, Bulletin, No. 312, High School Series No. 6.
54. Frailey, L. S., and Crain, C. M. "Correlation of Excellence in Different School Subjects Based on a Study of School Grades." *Journal of Educational Psychology*, March, 1914.
55. Holmes, W. H. *School Organization and the Individual Child, Grading and Special Schools.* The Davis Press, Worcester, Massachusetts, 1912.
56. Van Sickle, J. H., Witmer, L., and Ayres, L. P. *Provisions for Exceptional Children in the Public Schools.* United States Bureau of Education, Bulletin No. 14, 1911.

D. ON TEACHERS' MARKS AND MARKING SYSTEMS

57. Rugg, H. O. "Teachers' Marks and Marking Systems." *Journal of Educational Administration and Supervision*, February, 1915. Complete bibliography and summary of published literature to February, 1915.

IV. STUDIES OF PUBLIC SCHOOL COSTS AND BUSINESS MANAGEMENT

A. PUBLICATIONS RELATED TO CITY SCHOOL COSTS

(*a*) *Studies based on facts collected by question-blank methods*

58. Elliott, E. C. *Some Fiscal Aspects of Public Education.* Teachers College, Columbia University, Contributions to Education, No. 6, 1905.
58*a*. Monroe, W. S. *The Cost of Instruction in Kansas High Schools.* Emporia State Normal School, Bulletin No. 2, 1915.
59. Strayer, G. D. *City School Expenditures.* Teachers College, Columbia University, Contributions to Education, No. 5, 1905.

(*b*) *Studies based on facts collected personally from records of city systems*

60. Updegraff, H. *Study of Expenses of City School Systems.* United States Bureau of Education, Bulletin No. 5, 1912.
61. Hutchinson, J. H. *School Costs and School Accounting.* Teachers College, Columbia University, Contributions to Education, No. 62, 1914.

(*c*) *Reports of surveys of city school systems*

62. Clark, E. *Financing the Public Schools.* Cleveland Education Survey Monographs, Russell Sage Foundation, Division of Education, New York City.
63. Rugg, H. O. *Cost of Public Education in Grand Rapids, Michigan.* Chaps. XIV and XV of *Grand Rapids School Survey.* Board of Education, Grand Rapids, Michigan.
64. Rugg, H. O. *Public School Costs and Business Management in St. Louis.* (To be published, 1917, by Board of Education.)
65. Schroeder, H. H. *Cost of Public Education, Peoria, Illinois, 1915–16.* Board of Education, Peoria.

(*d*) *School reports as a means of presenting financial facts.*

66. Spaulding, F. E. 1912 and 1913, Newton, Massachusetts, School Reports. (Out of print.)
67. Spaulding, F. E. Three Monographs on School Finance in Minneapolis. Board of Education, Minneapolis, Minnesota:
    (*a*) *A Million a Year.*
    (*b*) *Financing the Minneapolis Schools.*
    (*c*) *The Price of Progress.*

(*e*) *General descriptive articles on financial practices of cities*

68. Baker, G. M. "Financial Practices in Cities and Towns below Twenty-five Thousand." *American School Board Journal*, October, November, December, 1916: January, February, March, May, June, 1917.
69. Clark, E. "The Indebtedness of City School Systems and Current School Expenditures." *American School Board Journal*, March, 1917, p. 17.

B. PUBLICATIONS RELATING TO THE STUDY OF BUSINESS MANAGEMENT OF THE PUBLIC CITY SCHOOLS

70. Byrne, J. T. Report of School Survey. Part IV: *The Business Management.* School Survey Committee, Denver, Colorado.
71. *Proceedings of National Association of School Accounting Officers*, published for the years 1913, 1914, 1915, 1916. Addresses reprinted in *American School Board Journal* for those years.
72. Shapleigh, F. E. "The Compensation of School Janitors." *American School Board Journal*, December, 1916, p. 23.
73. Hanus, P. H. "Town and City School Reports, more particularly Superintendents' Reports." *School and Society*, January 29, and February 5, 1916.

See also complete discussions in Nos. 63 and 64, Rugg, H. O.

C. PUBLICATIONS RELATING TO STATE AND COUNTY SCHOOL FINANCE

74. Cubberley, E. P. *School Funds and their Apportionment.* Teachers College, Columbia University, Contributions to Education, No. 2, 1915.
75. Swift, F. H. *A History of Public Permanent School Funds in the United States.* Henry Holt & Co., 1911.

76. Cubberley, E. P., and Elliott, E. C. *State and County School Administration.* The Macmillan Company, 1915.

77. Cubberley, E. P. *State and County Educational Reorganization.* The Macmillan Company, 1914.

78. MacDowell, T. L. *State* vs. *Local Control of Elementary Education.* (Finance.) United States Bureau of Education, Bulletin No. 22, 1915.

V. STUDIES OF CENTRAL ORGANIZATION AND ADMINISTRATION

5–11. Ayres, L. P. See Nos. 5–11.
Bobbitt, J. F. See No. 103 — Part I.

79. Chamberlain, A. H. *The Growth of Responsibility and Enlargement of Power of the City School Superintendent.* 158 pp. University of California Publications. *Education*, vol. III, No. 4, 1913.

80. Douglass, A. A. *The Junior High School.* Part III, Fifteenth Yearbook of the National Society for the Study of Education. 157 pp. The Public School Publishing Co., Bloomington, Illinois, 1917.

81. Shapleigh, F. E. "Commission Government and the Administration of City School Systems." *American School Board Journal*, November, 1915, p. 11.

82. Shapleigh, F. E. "School Administration in Non-Commission-Governed Cities." *American School Board Journal*, December, 1915, p. 11.

SUPERVISED STUDY

83. Breslich, E. R. *Supervised Study as a Means of Providing Supplementary Individual Instruction.* Thirteenth Yearbook of the National Society for the Study of Education, Part I, 1914.

84. Hall-Quest, A. L. *Supervised Study.* The Macmillan Company, 1916.

85. Minnich, J. H. "An Experiment in the Supervised Study of Mathematics." *School Review*, December, 1913, pp. 670–75.

86. Reavis, W. C. "Factors that Determine the Habits of Study in Grade Pupils." *Elementary School Teacher*, October, 1911, vol. XII, pp. 71–81.

GENERAL SUPERVISION OF INSTRUCTION

87. Twelfth Yearbook of National Society for the Study of Education. Part I: *The Supervision of City Schools*, 1913.
88. Twelfth Yearbook of National Society for the Study of Education. Part II: *The Supervision of Rural Schools*, 1913.
89. Jessup, W. A. *Social Factors Affecting Special Supervision.* Teachers College, Columbia University, Contributions to Education, No. 43, 1911.

VI. A BIBLIOGRAPHY OF SCHOOL SURVEYS

A. SURVEYS OF CITY SCHOOL SYSTEMS

90. Atlanta, Georgia. *Report of Survey of the Department of Education*, 1912. By New York Bureau of Municipal Research.

90*a*. Baltimore, Md. *Report of the Commission Appointed to Study the System of Education in the Public Schools of Baltimore.* U.S. Bureau of Education, Bulletin No. 4, 1911.

91. Blaine, Washington. *A Survey of the Blaine Public Schools.* By Lull, H. G., Millay, F. E., and Kruse, P. J. The Extension Division, University of Washington, Seattle, Washington, 1914.
92. Boise, Idaho. *Expert Survey of Public School System.* By Elliott, E. C., Judd, C. H., and Strayer, G. D. Board of Education, Boise, Idaho, 1913.
93. Boise, Idaho. *Special Report of the Boise Public Schools, 1915.*
94. Boston, Massachusetts. The Finance Commission of the City of Boston. *Report on the Boston School System.* City of Boston, Printing Department, 1911.
95. Boston, Massachusetts. *Report of a Study of Certain Phases of the Public School System of Boston, Massachusetts.* Sold by Teachers College, Columbia University, New York City, 1916.
96. Bridgeport, Connecticut. *Report of the Examination of the School System, 1915.* By Van Sickle, J. H.
97. Buffalo, New York. *Examination of the Public School System of the City of Buffalo.* By the Education Department of the State of New York. Albany. University of the State of New York, 1916.
98. Butte, Montana. *Report of a Survey of the School System of*

Butte, Montana. By Strayer, G. D., and others. Board of School Trustees, 1914.

99. Chicago, Illinois. *Report of the Educational Commission of the City of Chicago, 1897.*

100. Cleveland, Ohio. *The Cleveland Education Survey.* See Bibliography in Section on Course of Study. Published as 25 separate monographs. Ayres, L. P., Director. Russell Sage Foundation, New York City.

101. Dallas, Texas. *Report of the Public Schools of the City of Dallas, Texas.* Wilkinson Printing Co., Dallas, 1915.

102. Dansville, New York. *A Study — The Dansville High School.* By Foster, J. M. F. A. Owen Publishing Co., Dansville, New York.

103. Denver, Colorado. *Report of the School Survey of School District Number One in the City and County of Denver.* Part I: *General Organization and Management.* Part II: *The Work of the Schools.* Part III: *The Industrial Survey.* Part IV: *The Business Management.* Part V: *The Building Situation and Medical Inspection.* The School Survey Committee, Denver, Colorado, 1916.

104. East Orange, New Jersey. *Report of the Examination of the School System of East Orange, New Jersey, 1912.* By Moore, E. C. Issued by the Board of Education, 1912.

105. Grafton, West Virginia. *Report of the Survey of the Grafton City Schools.* By Deahl, J. N., Rosier, J., and Wilson, O. G. Department of Schools, Charleston, West Virginia, 1913.

106. Grand Junction, Colorado. *A Survey of the City Schools of Grand Junction, Colorado, District No. 1, Mesa County.* By Clapp, F. L., and others. The Daily News Press, 1916.

107. Grand Rapids, Michigan. *School Survey.* Grand Rapids, Michigan, 1916.

108. Greenwich, Connecticut. *The Book of the Educational Exhibit of Greenwich, Connecticut, 1912.* Conducted by the Russell Sage Foundation.

109. Hammond, Indiana. *Some Facts Concerning the People, Industries, and Schools of Hammond and a Suggested Program for Elementary, Industrial, Prevocational and Vocational Education.* By Leonard, R. J. Hammond, Indiana, Board of Education, 1915.

110. Leavenworth, Kansas. *Report of a Survey of the Public Schools of Leavenworth, Kansas.* Bureau of Educational Measurement

and Standards, Kansas State Normal School, Emporia, Kansas, 1915.

111. Minneapolis, Minnesota. *Report on the Survey of the Business Administration of the Minneapolis Public Schools.* By Bureau of Municipal Research of the Minneapolis Civic and Commerce Association, 1915.

112. Montclair, New Jersey. *Report on the Programme of Studies in the Public Schools of Montclair, New Jersey, 1911.* By Hanus, P. H.

113. Newburgh, New York. *The Newburgh Survey.* Department of Surveys and Exhibits, Russell Sage Foundation, 128 East Twenty-Third Street, New York City, 1913.

114. New York City. *Report of Committee on School Inquiry.* By Hanus, P. H., and others. Separate Reports reprinted as single volumes of World Book Company's School Efficiency Series, Yonkers, New York.

115. Oakland, California. *Report of a Survey of the Organization, Scope, and Finances of the Public System of Oakland, California.* By Cubberley, E. P. Board of Education, Bulletin No. 8, 1915.

116. Ogden, Utah. *Report of Ogden Public School Survey Commission.* By Deffenbaugh, W. S. State Department of Education, 1914.

117. Peoria, Illinois. *Cost of Public Education, 1915–1916, Peoria, Illinois.* By Schroeder, H. H. 1916.

118. Portland, Oregon. *The Portland Survey.* By Cubberley, E. P., and others. 1913.

119. Port Townsend, Washington. *A Survey of the Port Townsend Public Schools.* By Lull, H. G. University of Washington, University Extension Division, 1915.

120. Rockford, Illinois. *A Review of the Rockford Public Schools, 1915–1916.* Issued by the Board of Education, Rockford, Illinois, 1916.

121. Salt Lake City, Utah. *Report of a Survey of the School System of Salt Lake City, Utah.* By Cubberley, E. P., and others. Board of Education, Salt Lake City, 1915.

122. San Antonio, Texas. *The San Antonio Public School System.* By Bobbitt, J. F. The San Antonio School Board, 1915.

123. San Francisco, California. *Some Conditions in the Schools of San Francisco.* By Steinhart, Mrs. J. H., and others. School Survey Class, San Francisco, California, 1914.

124. Springfield, Illinois. *The Public Schools of Springfield, Illinois.* By Ayres, L. P. Division of Education, Russell Sage Foundation, New York City, 1914.

125. South Bend, Indiana. *Superintendent's Report of the School City of South Bend, Indiana, and School Survey.* By Department of Education of the University of Chicago, 1914.

126. St. Louis, Missouri. *Report of Survey of St. Louis School System.* (To be published, 1917, by Board of Education.)

127. Syracuse, New York. *Report of Investigations for the Associated Charities of Syracuse, New York.* Made by the Training School for Public Service. Conducted by the Bureau of Municipal Research, New York City, 1912.

128. Waterbury, Connecticut. *Help Your School Surveys.* By Brittain, H. L. New York Bureau of Municipal Research. Published together with report on instruction in St. Paul, Minnesota, by A. W. Farmer.

B. STATE SCHOOL SURVEYS

129. Connecticut. *Report of Education Commission,* in *Report of the Board of Education of the State of Connecticut,* 1909.

130. Colorado. *A General Survey of Public High-School Education in Colorado.* University of Colorado Bulletin, vol. xiv, No. 10, 1914.

131. Colorado. *Report of an Inquiry into the Administration and Support of the Colorado School System.* Department of Interior, United States Bureau of Education, Bulletin No. 5, 1917.

132. Illinois. *Illinois School Survey.* Coffman, L. D., Director. Published by J. A. Browne, Bloomington, Illinois, for the Illinois State Teachers' Association, 1917.

133. Iowa. *State Higher Educational Institutions of Iowa.* Commissioner of Education, Department of Interior, United States Bureau of Education, Bulletin No. 19, 1916. Washington, Government Printing Office, 1916.

134. Kansas. *Survey of Accredited High Schools and Professional Directory.* By Josselyn, H. W. Bulletin of the University of Kansas, vol. xv, No. 16, 1914.

135. Maryland. *Public Education in Maryland. A Report to the Maryland Educational Survey Commission.* General Education Board, New York, 1916.

136. Massachusetts. *Report of the Industrial Commission of the*

State of Massachusetts, 1907. Reprinted by Teachers College, Columbia University, New York City.

137. North Dakota. *Report of the Temporary Educational Commission to the Governor and Legislature of the State of North Dakota.* 1912.
138. North Dakota. *State Higher Educational Institutions of North Dakota.* Department of Interior, United States Bureau of Education, Bulletin No. 27, 1916. Washington, Government Printing Office, 1917.
139. Ohio. *Report of the Ohio State School Survey Commission.* By Campbell, M. E., Allendorf, W. L., and Thatcher, C. J. 1914.
140. United States. *A Comparative Study of Public School Systems in the Forty-eight States.* Russell Sage Foundation, Division of Education, 1912.
141. Vermont. *Education in Vermont.* The Carnegie Foundation for the Advancement of Teaching, Bulletin No. 7, 1914.
142. Virginia. *Report of the Virginia Education Commission to the General Assembly of the Commonwealth of Virginia.* 1912.
143. Washington. *A Survey of Educational Institutions of the State of Washington.* United States Bureau of Education, Bulletin No. 26, 1916.
144. Wyoming. *Educational Survey of Wyoming.* By Monahan, A. C., and Cook, K. M. United States Bureau of Education, Bulletin No. 29, 1916.

C. RURAL SCHOOL SURVEYS

145. Alabama, Three Counties. *An Educational Survey of Three Counties in Alabama.* Department of Education, Montgomery, Alabama, Bulletin No. 43, 1914.
146. Colorado. *The Rural and Village Schools of Colorado.* By Sargent, S. C. Colorado Agricultural College, Fort Collins, Colorado, 1914.
147. Indiana, Porter County. *Rural School Sanitation.* By Clark, T., Collins, G. L., and Treadway, W. L. Treasury Department, United States Public Health Service, 1916.
148. Maryland, Montgomery County. *An Educational Survey of a Suburban and Rural County.* By Morse, H. N., Eastman, F., and Monahan, A. C. United States Bureau of Education, Bulletin No. 32, 1913.
149. Missouri, Saline County. *A Study of the Rural Schools of*

Saline County, Missouri. By Elliff, J. D., and Jones, A. University of Missouri Bulletin, vol. 16, No. 22. University of Missouri, Columbia, Missouri, 1915.

150. *Reports in Westchester County, New York: A Study of Local School Conditions, 1912.* By Inglis, A. J.
151. Texas. *A Study of Rural Schools in Texas.* By White, E. V., and Davis, E. E. University of Texas, Austin, Texas, 1914.
152. Texas, Travis County. *A Study of Rural Schools in Travis County, Texas.* By Davis, E. E. University of Texas, Austin, Texas, 1916.
153. Virginia, Orange County. *Sanitary Survey of the Schools of Orange County, Virginia.* By Flannagan, R. K. United States Bureau of Education, Bulletin No. 17, 1914.

D. VOCATIONAL AND INDUSTRIAL SURVEYS

154. Denver, Colorado. *Report of the School Survey of School District Number One in the City and County of Denver.* Part III: *Vocational Education.* 1916.
155. *Minneapolis Survey for Vocational Education, Report of the.* National Society for the Promotion of Industrial Education, Bulletin No. 21, 1916.
156. New York City. *Seventeenth Annual Report of the City Superintendent of Schools, 1914–1915. Survey of the Gary Prevocational Schools.* Department of Education. City of New York.
157. Richmond, Virginia. *Vocational Education Survey of Richmond, Virginia.* United States Department of Labor, Bureau of Labor Statistics, 1916.
158. *A Survey of Manual, Domestic, and Vocational Training in the United States.* Hackett, W. E. Public Schools, Department of Practical Arts, Reading, Pennsylvania, 1914.

MISCELLANEOUS

159. *Some Foreign Educational Surveys.* By Mahoney, J. United States Bureau of Education, 1915.
160. *Community Action through Surveys.* By Harrison, S. M. Department of Surveys and Exhibits, Russell Sage Foundation, New York City, 1916.
161. Williams, J. H. *Reorganizing a County System of Schools.* United States Bureau of Education, Bulletin No. 16, 1916.

VII. REFERENCES CONTAINING DETAILED BIBLIOGRAPHIES OF VARIOUS PHASES OF SCHOOL ADMINISTRATION

162. Cubberley, E. P. *Public School Administration.* Houghton Mifflin Company, Boston, 1916.
163. Strayer, G. D., and Thorndike, E. L. *Educational Administration.* The Macmillan Company, New York, 1913.
13, 14, and 15. *Fourteenth, Fifteenth, and Sixteenth Yearbooks of National Society for the Study of Education.* See 13, 14, 15.
164. Holmes, H. W., and others. *A Descriptive Bibliography of Measurement in Elementary Subjects.* Harvard University Press, Cambridge, 1917.
41. Holley, C. E. *The Relationship between Persistence in School and Home Conditions.* See 41.
80. Douglass, A. A. *The Junior High School.* See 80.
84. Hall-Quest, A. L. *Supervised Study.* See 84.
165. Rugg, H. O. "Summary of the Literature on Public School Costs and Business Management." *Elementary School Journal,* April, 1917.
57. Rugg, H. O. *Teachers' Marks and Marking Systems.* See 57.

APPENDIX A

SELECTED BIBLIOGRAPHY ON STATISTICAL METHODS

A. HISTORY OF STATISTICS

Meitzen, A. "History, Theory and Technique of Statistics." Annals, American Academy of Political and Social Science. Philadelphia, 1891, Part I. Translation by R. P. Falkner.

B. GENERAL AND NON-MATHEMATICAL BOOKS

Elderton, W. P., and E. M. *Primer of Statistics.* A. & C. Black, London, 1910.

King, W. I. *The Elements of Statistical Method.* The Macmillan Company, New York, 1915.

Zizek, F. *Statistical Averages.* Translated by W. M. Persons. Henry Holt & Co., New York, 1913.

C. GENERAL BOOKS INVOLVING NO CALCULUS

Bowley, A. L. *Elements of Statistics.* P. S. King & Son, London; Charles Scribner's Sons, New York, 1907.

Bowley, A. L. *An Elementary Manual of Statistics.* McDonald & Evans, London, 1910.

Yule, G. U. *An Introduction to the Theory of Statistics.* C. Griffin & Co., London, 1912.

D. BOOKS AND PAMPHLETS SUMMARIZING STATISTICAL METHODS AND GIVING TABLES

Davenport, C. B. *Statistical Methods.* (Especially adapted to the study of biological variation). Summarizes all mathematical formulæ with brief descriptions of methods. Complete and useful tables, containing probability integrals, Beta and Gamma functions, logarithms, etc. J. Wiley & Sons, New York, 1904.

Kelley, T. L. *Tables: To Facilitate the Calculation of Partial Coefficients of Correlation and Regression Coefficients.* University of Texas, Bulletin No. 27, 1916. Austin, Texas.

Rietz, H. L. Appendix to E. Davenport's *Principles of Breeding.* Ginn & Co., Boston, 1907.

Rietz, H. L. Bulletins 119 and 148 of the University of Illinois Agricultural Experimentation Station. Summaries of statistical methods involving little advanced mathematics.

Whipple, G. M. *Manual of Mental and Physical Tests.* Vol. 1, *Simpler Processes,* chap. III. A brief summary of the statistical treatment of measures as applied to problems of educational psychology. Warwick & York, Baltimore, 2d edition, 1914.

E. CALCULATING TABLES

Barlow's *Tables of Squares, Cubes, Square-roots, Cube-roots and Reciprocals of all Integers, Numbers up to 10,000.* E. Spon, New York.

Cotsworth, M. B. *The Direct Calculator. Series O. R.* M. B. Cotsworth, Holgate, York, England.

Crelle, A. L. *Rechentafeln.* G. Reiner, Berlin, new edition, 1907. (Cotsworth and Crelle give products to 1000 by 1000.)

Elderton, W. P. *Tables of Powers of Natural Numbers and of the Sums of Powers of the Natural Numbers from 1 to 100. Biometrika,* vol. II, p. 474.

Peters, J. *Neue Rechentafeln für Multiplikation und Division.* G. Reimer, Berlin.

F. BOOKS ADAPTING STATISTICAL METHODS TO EDUCATIONAL AND PSYCHOLOGICAL INVESTIGATION

Brown, *The Essentials of Mental Measurement.* Cambridge University, 1911.

Thorndike, E. L. *An Introduction to the Theory of Mental and Social Measurements.* Teachers College, Columbia University, New York, 1913.

G. BOOKS ON GRAPHIC METHODS

Brinton, W. C. *Graphic Methods for Presenting Facts.* Engineering Magazine Company, New York, 1914.

H. FOR SUMMARY OF MATHEMATICAL THEORY UNDERLYING CORRELATION AND TYPE CURVES OF DISTRIBUTION, *see*

Elderton, W. P. *Frequency Curves and Correlation.* C. & E. Layton, London, 1906.

I. FOR A TREATMENT OF LEAST SQUARES, *see*

Merriman, M. *A Textbook on the Method of Least Squares.* J. Wiley and Sons, New York.

J. THE ORIGINAL CONTRIBUTIONS TO THE MATHEMATICAL FOUNDATION OF STATISTICAL METHODS WILL BE FOUND PRINCIPALLY IN

Journal of the Royal Statistical Society; Philosophical Transactions of the Royal Society; Biometrika; Draper's Company *Research; Memoirs; Philosophical Magazine.*

K. A COMPLETE BIBLIOGRAPHY OF THE ORIGINAL MEMOIRS WILL BE FOUND IN

Yule, G. U. *Introduction to the Theory of Statistics.* C. Griffin & Co., London, 1912.

L. PROBLEM BOOKS IN EDUCATIONAL STATISTICS

Rugg, H. O. *Illustrative Problems in Educational Statistics.* Published by the author, University of Chicago Press, 1916.

APPENDIX B

SUMMARY OF FORMULÆ AND SYMBOLS USED IN THE TEXT

CHAPTERS IV AND V

f = frequency of measures
m = a measure
N = total number of cases
d = deviation (used for deviation in units of class-intervals)
M = arithmetical mean = $\frac{\Sigma fm}{N}$
Md = median (= Q_2)
M_o = mode
c = correction applied to assumed mean to obtain true mean
H = Harmonic mean. $\frac{1}{H} = \frac{1}{N} \Sigma \left(\frac{1}{m}\right)$
r_{ave} = average rate
t_{ave} = average time
M_G = geometric mean = $\sqrt[n]{m_1 . m_2 . m_3 . m_4 . \ldots\ldots\ldots\ldots m_n}$

CHAPTER VI

Q_1 = first or lower quartile point
Q_3 = third or upper " "
Q = quartile deviation or semi-interquartile-range = $\frac{Q_3 - Q_1}{2}$
σ = standard deviation (sometimes represented $S.D.$ or ϵ) = $\sqrt{\frac{\Sigma fd^2}{N}}$
$P.E.$ = Probable error = $.6745\sigma$
$M.D.$ = Mean deviation = $\frac{\Sigma fd}{N}$
V = Pearson's coefficient of variation = $\frac{100\sigma}{M}$, or $\frac{100\ M.D.}{Md}$
Skewness = $\frac{\text{mean} - \text{mode}}{\sigma}$ or $\frac{3\ (\text{mean} - \text{median})}{\sigma} = \frac{Q_1 + Q_3 - 2Md}{Q}$

CHAPTER VII

^{r} or $r!$ = product of all integers from 1 to r, $= 1 \cdot 2 \cdot 3 \cdot 4 \ldots\ldots r$

$n^Pr = n(n-1)\ (n-2) \ldots\ldots (n-r+1)$

$$n^Cr = \frac{n(n-1)\ (n-2) \ldots\ldots (n-r+1)}{\underline{|r}}$$

$$(P+Q)^n = P^n + nP^{n-1}q + \frac{n(n-1)}{1 \cdot 2} P^{n-2}\, q^2 + \ldots \frac{n(n-1)\ (n-2)}{n-r+1} p^{n-r} q^r$$

$$\left(\frac{1}{2}+\frac{1}{2}\right)^n$$

$$= \left(\frac{1}{2}\right)^n + n\left(\frac{1}{2}\right)^n + \frac{n(n-1)}{1 \cdot 2}\left(\frac{1}{2}\right)^n + \frac{n(n-1)\,(n-2)}{1 \cdot 2 \cdot 3}\left(\frac{1}{2}\right) + \ldots$$

$y = y_o e^{\frac{-x^2}{2\sigma^2}}$ = equation of "normal" probability curve

CHAPTER VIII

$y_o = \dfrac{N}{\sigma\sqrt{2\pi}}$ = equation of mean ordinate of "normal" probability curve

σ_S = standard error of sampling $= \sqrt{\dfrac{f\,(N-f)}{N}}$

σ_M = standard deviation of a mean $= \dfrac{\sigma_{distribution}}{\sqrt{N}}$

$P.E._M$ = probable error $= .6745\ \dfrac{\sigma_{distribution}}{\sqrt{N}}$

σ_σ = standard deviation of a standard deviation $= \dfrac{\sigma_{distribution}}{\sqrt{2N}}$

$P.E._\sigma$ = probable error of a standard deviation

$= .6745\ \dfrac{\sigma_{distribution}}{\sqrt{2N}}$

σ_r = standard deviation of a coefficient of correlation $= \dfrac{1-r^2}{\sqrt{N}}$

$P.E._r$ = probable error of a coefficient of correlation

$= .6745\ \dfrac{1-r^2}{\sqrt{N}}$

CHAPTER IX

$y_1 - \bar{y} = r \frac{\sigma_y}{\sigma_x} (x_1 - \bar{x})$, = the equation of straight line of regression; x_1 and y_1 in terms of actual values of measures.

$y = r \frac{\sigma_y}{\sigma_x} x$, and $x = r \frac{\sigma_x}{\sigma_y} y$, are equations for straight line regression with x and y expressed as deviations of particular x and y measures from their mean values; *i.e.*, $y = (y_1 - \bar{y})$ $x = (x_1 - \bar{x})$

r = coefficient of correlation = $\frac{\Sigma xy}{N\sigma_x\sigma_y}$

$r = \frac{\frac{\Sigma x'y'}{N} - c_x c_y}{\sigma_x \sigma_y}$ = formula for short method of computing r

$b_1 = r \frac{\sigma_y}{\sigma_x}$ = regression coefficient of y on x

$b_2 = r \frac{\sigma_x}{\sigma_y}$ = regression coefficient of x on y

$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$ = formula for computing r without tabulation of correlation table

$\eta = \frac{\Sigma}{\sigma_y} = \frac{\sqrt{\frac{S[n_x(\bar{y}_x - \bar{y})^2]}{N}}}{\sigma_y}$ = the correlation ratio.

$r = 2 \sin (\frac{\pi}{6} \rho)$ = Pearson's formula for correlation of grades in which $\rho = 1 - \frac{6\, \Sigma D^2}{N(N^2 - 1)}$

$r = 2 \cos \frac{\pi}{3} (1 - R) - 1$ = Pearson's formula for correlation by grades using the positive signed differences only;

in which $R = 1 - \frac{6 \Sigma\, g}{N^2 - 1}$. R is known as Spearman's Foot-Rule for measuring correlation.

$r = \cos \frac{\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}} \pi$ = Pearson's formula for measuring correlation for fourfold tables.

$r = \cos \frac{}{L + U} \pi$ = Sheppard's formula for measuring correlation for fourfold tables.

$C = \sqrt{\frac{\chi^2}{N + \chi^2}}$ or $C = \sqrt{\frac{S - N}{S}}$ = Pearson's formula for coefficient of *mean-square-contingency*

$\chi^2 = \left\{ \frac{(n_{rc})^2}{\frac{n_r \quad n_c}{N}} \right\} - N$ = Pearson's *square-contingency*.

SYMBOLS AND NUMERICAL VALUES FOR CERTAIN CONSTANTS USED IN SCHOOL RESEARCH

π	Ratio of Circumference to Diameter	3.14159
$\frac{1}{\pi}$	Reciprocal of π	.31831
$\sqrt{2\pi}$	Square root of 2π	2.506628
e	Base of Napierian or hyperbolic logarithms	2.71828

APPENDIX C

TABLES TO FACILITATE COMPUTATION

Table I. Natural Sines and Cosines

′	0°		1°		2°		3°		4°		′
	sin	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	
0	0000	1.000	0175	9998	0349	9994	0523	9986	0698	9976	60
5	0015	1.000	0189	9998	0364	9993	0538	9986	0712	9975	55
10	0029	1.000	0204	9998	0378	9993	0552	9985	0727	9974	50
15	0044	1.000	0218	9998	0393	9992	0567	9984	0741	9973	45
20	0058	1.000	0233	9997	0407	9992	0581	9983	0756	9971	40
25	0073	1.000	0247	9997	0422	9991	0596	9982	0770	9970	35
30	0087	1.000	0262	9997	0436	9990	0610	9981	0785	9969	30
35	0102	9999	0276	9996	0451	9990	0625	9980	0799	9968	25
40	0116	9999	0291	9996	0465	9989	0640	9980	0814	9967	20
45	0131	9999	0305	9995	0480	9988	0654	9979	0828	9966	15
50	0145	9999	0320	9995	0494	9988	0669	9978	0843	9964	10
55	0160	9999	0334	9994	0509	9987	0683	9977	0857	9963	5
60	0175	9999	0349	9994	0523	9986	0698	9976	0872	9962	0
	cos	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	
′	89°		88°		87°		86°		85°		′

′	5°		6°		7°		8°		9°		′
	sin	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	
0	0872	9962	1045	9945	1219	9925	1392	9903	1564	9877	60
5	0886	9961	1060	9944	1233	9924	1406	9901	1579	9875	55
10	0901	9959	1074	9942	1248	9922	1421	9899	1593	9872	50
15	0915	9958	1089	9941	1262	9920	1435	9897	1607	9870	45
20	0929	9957	1103	9939	1276	9918	1449	9894	1622	9868	40
25	0944	9955	1118	9937	1291	9916	1464	9892	1636	9865	35
30	0958	9954	1132	9936	1305	9914	1478	9890	1650	9863	30
35	0973	9953	1146	9934	1320	9913	1492	9888	1665	9860	25
40	0987	9951	1161	9932	1334	9911	1507	9886	1679	9858	20
45	1002	9950	1175	9931	1349	9909	1521	9884	1693	9856	15
50	1016	9948	1190	9929	1363	9907	1536	9881	1708	9853	10
55	1031	9947	1204	9927	1377	9905	1550	9879	1772	9851	5
60	1045	9945	1219	9925	1392	9903	1564	9877	1736	9848	0
	cos	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	
′	84°		83°		82°		81°		80°		′

Table I (*continued*)

′	10°		11°		12°		13°		14°		′
	sin	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	
0	1736	9848	1908	9816	2079	9781	2250	9744	2419	9703	60
5	1751	9846	1922	9813	2093	9778	2264	9740	2433	9699	55
10	1765	9843	1937	9811	2108	9775	2278	9737	2447	9696	50
15	1779	9840	1951	9808	2122	9772	2292	9734	2462	9692	45
20	1794	9838	1965	9805	2136	9769	2306	9730	2476	9689	40
25	1808	9835	1979	9802	2150	9766	2320	9727	2490	9685	35
30	1822	9833	1994	9799	2164	9763	2334	9724	2504	9681	30
35	1837	9830	2008	9796	2179	9760	2349	9720	2518	9678	25
40	1851	9827	2022	9793	2193	9757	2363	9717	2532	9674	20
45	1865	9825	2036	9790	2207	9753	2377	9713	2546	9670	15
50	1880	9822	2051	9787	2221	9750	2391	9710	2560	9667	10
55	1894	9819	2065	9784	2235	9747	2405	9706	2574	9663	5
60	1908	9816	2079	9781	2250	9744	2419	9703	2588	9659	0
	cos	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	
′	79°		78°		77°		76°		75°		

′	15°		16°		17°		18°		19°		′
	sin	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	
0	2588	9659	2756	9613	2924	9563	3090	9511	3256	9455	60
5	2602	9655	2770	9609	2938	9559	3104	9506	3269	9450	55
10	2616	9652	2784	9605	2952	9555	3118	9502	3283	9446	50
15	2630	9648	2798	9600	2965	9550	3132	9497	3297	9441	45
20	2644	9644	2812	9596	2979	9546	3145	9492	3311	9436	40
25	2659	9640	2826	9592	2993	9542	3159	9488	3324	9431	35
30	2672	9636	2840	9588	3007	9537	3173	9483	3338	9426	30
35	2686	9632	2854	9584	3021	9533	3187	9479	3352	9422	25
40	2700	9628	2868	9580	3035	9528	3201	9474	3365	9417	20
45	2714	9625	2882	9576	3048	9524	3214	9469	3379	9412	15
50	2728	9621	2896	9572	3062	9520	3228	9456	3393	9407	10
55	2742	9617	2910	9567	3076	9515	3242	9460	3407	9402	5
60	2756	9613	2924	9563	3090	9511	3256	9455	3420	9397	0
	cos	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	
′	74°		73°		72o		71°		70°		′

TABLE I (*continued*)

′	20°		21°		22°		23°		24°		′
	sin	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	
0	3420	9397	3584	9336	3746	9272	3907	9205	4067	9135	60
5	3434	9392	3597	9331	3760	9266	3921	9199	4081	9130	55
10	3448	9387	3611	9325	3773	9261	3934	9194	4094	9124	50
15	3461	9382	3624	9320	3786	9255	3947	9188	4107	9118	45
20	3475	9377	3638	9315	3800	9250	3961	9182	4120	9112	40
25	3488	9372	3651	9309	3313	9244	3974	9176	4134	9106	35
30	3502	9367	3665	9304	3827	9239	3987	9171	4147	9100	30
35	3516	9362	3679	9299	3840	9233	4001	9165	4160	9094	25
40	3529	9356	3692	9293	3854	9228	4014	9159	4173	9088	20
45	3543	9351	3706	9288	3867	9222	4027	9153	4187	9081	15
50	3557	9346	3719	9283	3881	9216	4041	9147	4200	9075	10
55	3570	9341	3733	9277	3894	9211	4054	9141	4313	9069	5
60	3584	9336	3746	9272	3907	9205	4037	9135	4226	9063	0
	cos	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	
′	69°		68°		67°		66°		65°		′

′	25°		26°		27°		28°		29°		′
	sin	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	
0	4226	9063	4384	8988	4540	8910	4695	8829	4848	8746	60
5	4239	9057	4397	8982	4553	8903	4708	8823	4861	8739	55
10	4253	9051	4410	8975	4566	8897	4720	8816	4874	8732	50
15	4266	9045	4423	8969	4579	8890	4733	8809	4886	8725	45
20	4279	9038	4436	8962	4592	8884	4746	8802	4899	8718	40
25	4292	9032	4449	8956	4605	8877	4759	8795	4912	8711	35
30	4305	9026	4462	8949	4617	8870	4772	8788	4924	8704	30
35	4318	9020	4475	8943	4630	8863	4784	8781	4937	8696	25
40	4331	9013	4488	8936	4643	8857	4797	8774	4950	8689	20
45	4344	9007	4501	8930	4656	8850	4810	8767	4962	8682	15
50	4358	9001	4514	8923	4669	8843	4823	8760	4975	8675	10
55	4371	8994	4527	8917	4682	8836	4835	8753	4987	8668	5
60	4384	8988	4540	8910	4695	8829	4848	8746	5000	8660	0
	cos	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	
′	64°		63°		62°		61°		60°		′

TABLE I (*continued*)

′	30°		31°		32°		33°		34°		′
	sin	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	
0	5000	8660	5150	8572	5299	8480	5446	8387	5592	8290	60
5	5013	8653	5163	8564	5312	8473	5459	8379	5604	8282	55
10	5025	8646	5175	8557	5324	8486	5471	8371	5616	8274	50
15	5038	8638	5188	8549	5336	8457	5483	8363	5628	8266	45
20	5050	8631	5200	8542	5348	8450	5495	8355	5640	8258	40
25	5963	8624	5213	8534	5361	8442	5507	8347	5652	8249	35
30	5075	8616	5225	8526	5373	8434	5519	8339	5664	8241	30
35	5088	8609	5237	8519	5385	8426	5531	8331	5676	8233	25
40	5100	8601	5250	8511	5398	8418	5544	8323	5688	8225	20
45	5113	8594	5262	8504	5410	8410	5556	8315	5700	8216	15
50	5125	8587	5275	8496	5422	8403	5568	8307	5712	8208	10
55	5138	8579	5287	8488	5434	8395	5580	8299	5724	8200	5
60	5150	8572	5299	8480	5446	8387	5592	8290	5736	8192	0
	cos	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	
′	59°		58°		57°		56°		55°		′

′	35°		36°		37°		38		39°		′
	sin	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	
0	5736	8192	5878	8090	6018	7986	6157	7880	6293	7771	60
5	5748	8183	5890	8082	6030	7978	6168	7871	6305	7762	55
10	5760	8175	5901	8073	6041	7969	6180	7862	6316	7753	50
15	5771	8166	5913	8064	6053	7960	6191	7853	6327	7744	45
20	5783	8158	5925	8056	6065	7951	6202	7844	6338	7735	40
25	5795	8150	5937	8047	6076	7942	6214	7835	6350	7725	35
30	5807	8141	5948	8039	6088	7934	6225	7826	6361	7716	30
35	5819	8133	5960	8030	6099	7925	6237	7817	6372	7707	25
40	5831	8124	5972	8021	6111	7916	6248	7808	6383	7698	20
45	5842	8116	5983	8013	6122	7907	6259	7799	6394	7688	15
50	5854	8107	5995	8004	6134	7898	6271	7790	6406	7679	10
55	5866	8099	6007	7995	6145	7889	6282	7781	6417	7670	5
60	5878	8090	6018	7986	6157	7880	6293	7771	6428	7660	0
	cos	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	
′	54°		53°		52°		51°		50°		′

Table I (*continued*)

′	40°		41°		42°		43°		44°		′
	sin	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	
0	6428	7660	6561	7547	6691	7431	6820	7314	6947	7193	60
5	6439	7651	6572	7538	6702	7422	6831	7304	6957	7183	55
10	6450	7642	6583	7528	6713	7412	6841	7294	6967	7173	50
15	6461	7632	6593	7518	6724	7402	6852	7284	6978	7163	45
20	6472	7623	6604	7509	6734	7392	6862	7274	6988	7153	40
25	6483	7613	6615	7499	6745	7383	6873	7264	6999	7143	35
30	6494	7604	6626	7490	6756	7373	6884	7254	7009	7133	30
35	6506	7595	6637	7480	6767	7363	6894	7244	7019	7122	25
40	6517	7585	6648	7470	6777	7253	6905	7234	7030	7112	20
45	6528	7576	6659	7461	6788	7343	6915	7224	7040	7102	15
50	6539	7566	6670	7451	6799	7333	6926	7214	7050	7092	10
55	6550	7557	6680	7441	6809	7323	6936	7203	7061	7081	5
60	6561	7547	6691	7431	6820	7314	6947	7193	7071	7071	0
	cos	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	*cos*	*sin*	
′	49°		48°		47°		46°		45°		′

TABLE II

Ordinates of the normal probability curve expressed as fractional parts of the mean ordinate y_o. Each ordinate is erected at a given distance from the mean. The height of the ordinate erected at the mean can be computed from,

$$y_o = \frac{N}{\sigma\sqrt{2\pi}} = \frac{N}{2.5066\,\sigma}$$

The corresponding height of any other ordinate can be read from the table by assigning the distance that the ordinate is from the mean (x). Distances on x are measured as fractional parts of σ. Thus the height of an ordinate at a distance from the mean of $.7\sigma$ will be $.78270\ y_o$; the height of an ordinate at $2.15\ \sigma$ from the mean will be $.09914\ y_o$, etc.

x/σ	0	1	2	3	4	5	6	7	8	9
0.0	100000	99995	99980	99955	99920	99875	99820	99755	99685	99596
0.1	99501	99396	99283	99158	99025	98881	98728	98565	98393	98211
0.2	98020	97819	97609	97390	97161	96923	96676	96420	96156	95882
0.3	95600	95309	95010	94702	94387	94055	93723	93382	93024	92677
0.4	92312	91399	91558	91169	90774	90871	89961	89543	89119	88688
0.5	88250	87805	87353	86896	86432	85962	85488	85006	84519	84060
0.6	83527	83023	82514	82010	81481	80957	80429	79896	79359	78817
0.7	78270	77721	77167	76610	76048	75484	74916	74342	73769	73193
0.8	72615	72033	71448	70861	70272	69681	69087	68493	67896	67298
0.9	66689	66097	65494	64891	64287	63683	63077	62472	61865	61259
1.0	60653	60047	59440	58834	58228	57623	57017	56414	55810	55209
1.1	54607	54007	53409	52812	52214	51620	51027	50437	49848	49260
1.2	48675	48092	47511	46933	46357	45783	45212	44644	44078	43516
1.3	42956	42399	41845	41294	40747	40202	39661	39123	38569	38058
1.4	37531	37007	36487	35971	35459	34950	34445	33944	33447	32954
1.5	32465	31980	31500	31023	30550	30082	29618	29158	28702	28251
1.6	27804	27361	26923	26489	26059	25634	25213	24797	24385	23978
1.7	23575	23176	22782	22392	22008	21627	21251	20879	20511	20148
1.8	19790	19436	19086	18741	18400	18064	17732	17404	17081	16762
1.9	16448	16137	15831	15530	15232	14939	14650	14364	14083	13806
2.0	13534	13265	13000	12740	12483	12230	11981	11737	11496	11259
2.1	11025	10795	10570	10347	10129	09914	09702	09495	09290	09090
2.2	08892	08698	08507	08320	08136	07956	07778	07604	07433	07265
2.3	07100	06939	06780	06624	06471	06321	06174	06029	05888	05750
2.4	05614	05481	05350	05222	05096	04973	04852	04734	04618	04505
2.5	04394	04285	04179	04074	03972	03873	03775	03680	03586	03494
2.6	03405	03317	03232	03148	03066	02986	02908	02831	02757	02684
2.7	02612	02542	02474	02408	02343	02280	02218	02157	02098	02040
2.8	01984	01929	01876	01823	01772	01723	01674	01627	01581	01536
2.9	01492	01449	01408	01367	01328	01288	01252	01215	01179	01145
3.0	01111	00819	00598	00432	00309	00219	00153	00106	00073	00050
4.0	00034	00022	00015	00010	00006	00004	00003	00002	00001	00001
5.0	00000									

TABLE III

Fractional parts of the total area (10,000) under the normal probability curve, corresponding to distances on the baseline between the mean and successive points of division laid off from the mean. Distances are measured in units of the standard deviation, σ. To illustrate, the table is read as follows: between the mean ordinate, y_o, and any ordinate erected at a distance from it of, say, $.8\sigma$ $\left(i.e., \frac{x}{\sigma} = .8 \right)$, is included 28.81 per cent of the entire area.

x/σ	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0000	0040	0080	0120	0159	0199	0239	0279	0319	0359
0.1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
0.2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0.3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0.4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0.5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0.6	2257	2291	2324	2357	2389	2422	2454	2486	2518	2549
0.7	2580	2612	2642	2673	2704	2734	2764	2794	2823	2852
0.8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0.9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1.0	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1.1	3643	3665	3686	3718	3729	3749	3770	3790	3810	3830
1.2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1.3	4032	4049	4066	4083	4099	4115	4131	4147	4162	4177
1.4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1.5	4332	4345	4357	4370	4382	4394	4406	4418	4430	4441
1.6	4452	4463	4474	4485	4495	4505	4515	4525	4535	4545
1.7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1.8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1.9	4713	4719	4726	4832	4738	4744	4750	4758	4762	4767
2.0	4773	4778	4783	4788	4793	4798	4803	4808	4812	4817
2.1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2.2	4861	4865	4868	4871	4875	4878	4881	4884	4887	4890
2.3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2.4	4918	4920	4922	4925	4927	4929	4931	4932	4934	4936
2.5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2.6	4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
2.7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2.8	4974	4975	4976	4977	4977	4978	4879	4980	4980	4981
2.9	4981	4982	4983	4984	4984	4984	4985	4985	4986	4986

TABLE III (*continued*)

x/σ	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
3.0	4986.5	4987	4987	4988	4988	4988	4989	4989	4989	4990
3.1	4990.3	4991	4991	4991	4992	4992	4992	4992	4993	4993
3.2	4993.129									
3.3	4995.166									
3.4	4996.631									
3.5	4997.674									
3.6	4998.409									
3.7	4998.922									
3.8	4999.277									
3.9	4999.519									
4.0	4999.683									
4.5	4999.966									
5.0	4999.997133									

TABLE IV

Fractional parts of the total area (10,000) under the normal probability curve, corresponding to distances on the base line between the mean and successive points of division laid off from the mean. Distances are measured in units of the Probable Error (*P.E.*). To illustrate, the table is read as follows: between the mean ordinate, y_o, and any ordinate erected at a distance from it of, say, 1.4 *P.E.* is $\left(i.e., \frac{x}{P.E.} = 1.4 \right)$ included 32.75 per cent of the entire area.

$\frac{x}{P.E.}$	*.00*	*.05*	$\frac{x}{P.E.}$	*.00*	*.05*	$\frac{x}{P.E.}$	*.00*	*.05*	$\frac{x}{P.E.}$	*.00*	*.05*
0	0000	0135	1.5	3441	3521	3.0	4785	4802	4.5	4988	4989
.1	0269	0403	1.6	3597	3671	3.1	4817	4831	4.6	4990	4991
.2	0536	0670	1.7	3742	3811	3.2	4845	4858	4.7	4992	4993
.3	0802	0933	1.8	3896	3939	3.3	4870	4881	4.8	4994	4994.6
.4	1063	1193	1.9	4000	4057	3.4	4891	4900	4.9	4995.2	4995.7
.5	1321	1447	2.0	4113	4166	3.5	4909	4917	5.0	4996.2	4996.6
.6	1571	1695	2.1	4217	4265	3.6	4924	4931	5.1	4997.1	4997.4
.7	1816	1935	2.2	4311	4354	3.7	4937	4943	5.2	4997.7	4998.0
.8	2053	2168	2.3	4396	4435	3.8	4948	4953	5.3	4998.2	4998.4
.9	2291	2392	2.4	4472	4508	3.9	4957	4961	5.4	4998.6	4998.8
1.0	2500	2606	2.5	4541	4573	4.0	4965	4968	5.5	4999.0	4999.1
1.1	2709	2810	2.6	4602	4631	4.1	4971	4974	5.6	4999.2	4999.3
1.2	2908	3004	2.7	4657	4682	4.2	4977	4979	5.7	4999.4	4999.5
1.3	3097	3188	2.8	4705	4727	4.3	4981	4983	5.8	4999.55	4999.6
1.4	3275	3360	2.9	4748	4767	4.4	4985	4987	5.9	4999.65	4999.7

TABLE V

Percentile scores to be assigned to test problems or questions which correspond to various percentages of pupils who fail to solve problems or questions correctly. Table is based upon area of normal probability curve, *assuming base line to be broken off at* $\pm 2.5\sigma$. Scholastic abilities are assumed to fit the probability curve and percentages of pupils who solve various problems correspond to percentages of area under the curve from the 0 point to a point on the base line. This point on base line, measured in units of σ, is transformed into percentile scores by setting 0 at -2.5σ, 50 at the mean, and 100 at $+2.5\sigma$.

Example: A problem failed by 22 per cent of a large number of pupils is scored 35; one failed by 98.35 per cent is scored 90; etc.

Per cent failing	*Distance* $\frac{x}{\sigma}$	*Percentile score*	*Per cent failing*	*Distance* $\frac{x}{\sigma}$	*Percentile score*	*Per cent failing*	*Distance* $\frac{x}{\sigma}$	*Percentile score*	*Per cent failing*	*Distance* $\frac{x}{\sigma}$	*Percentile score*
.02	.01		.73	.29		2.12	.58		4.43	.86	
.04	.02		.77	.30	6	2.19	.59		4.53	.87	
.06	.03		.81	.31		2.25	.60	12	4.64	.88	
.07	.04		.84	.32		2.32	.61		4.75	.89	
.09	.05	1	.88	.33		2.39	.62		4.86	.90	18
.11	.06		.92	.34		2.45	.63		4.97	.91	
.13	.07		.96	.35	7	2.52	.64		5.08	.92	
.16	.08		1.00	.36		2.60	.65	13	5.20	.93	
.18	.09		1.04	.37		2.67	.66		5.32	.94	
.20	.10	2	1.08	.38		2.74	.67		5.44	.95	19
.22	.11		1.12	.39		2.82	.68		5.55	.96	
.25	.12		1.17	.40	8	2.89	.69		5.68	.97	
.27	.13		1.21	.41		2.97	.70	14	5.81	.98	
.29	.14		1.26	.42		3.05	.71		5.93	.99	
.32	.15	3	1.30	.43		3.13	.72		6.06	1.00	20
.34	.16		1.35	.44		3.22	.73		6.19	1.01	
.37	.17		1.40	.45	9	3.30	.74		6.32	1.02	
.40	.18		1.45	.46		3.39	.75	15	6.46	1.03	
.42	.19		1.50	.47		3.47	.76		6.59	1.04	
.45	.20	4	1.60	.49		3.57	.77		6.73	1.05	21
.48	.21		1.65	.50	10	3.65	.78		6.87	1.06	
.51	.22		1.71	.51		3.74	.79		7.02	1.07	
.54	.23		1.76	.52		3.84	.80	16	7.16	1.08	
.57	.24		1.80	.53		3.93	.81		7.31	1.09	
.60	.25	5	1.88	.54		4.03	.82		7.46	1.10	22
.63	.26		1.94	.55	11	4.13	.83		7.61	1.11	
.67	.27		2.00	.56		4.23	.84		7.76	1.12	
.70	.28		2.06	.57		4.33	.85	17	7.91	1.13	

TABLE V (*continued*)

Per cent failing	*Distance* $\frac{x}{\sigma}$	*Percentile score*	*Per cent failing*	*Distance* $\frac{x}{\sigma}$	*Percentile score*	*Per cent failing*	*Distance* $\frac{x}{\sigma}$	*Percentile score*
8.07	1.14		17.00	1.57		30.23	2.00	40
8.23	1.15	23	17.26	1.58		30.59	2.01	
8.39	1.16		17.52	1.59		30.94	2.02	
8.55	1.17		17.79	1.60	32	31.30	2.03	
8.72	1.18		18.05	1.61		31.66	2.04	
8.89	1.19		18.32	1.62		32.02	2.05	41
9.06	1.20	24	18.60	1.63		32.38	2.06	
9.23	1.21		18.87	1.64		32.74	2.07	
9.46	1.22		19.15	1.65	33	33.10	2.08	
9.58	1.23		19.43	1.66		33.47	2.09	
9.76	1.24		19.71	1.67		33.84	2.10	42
9.94	1.25	25	19.99	1.68		34.21	2.11	
10.13	1.26		20.28	1.69		34.58	2.12	
10.31	1.27		20.57	1.70	34	34.95	2.13	
10.50	1.28		20.86	1.71		35.32	2.14	
10.69	1.29		21.15	1.72		35.70	2.15	43
10.89	1.30	26	21.44	1.73		36.07	2.16	
11.08	1.31		21.74	1.74		36.45	2.17	
11.28	1.32		22.04	1.75	35	36.83	2.18	
11.48	1.33		22.34	1.76		37.21	2.19	
11.68	1.34		22.65	1.77		37.59	2.20	44
11.89	1.35	27	22.96	1.78		37.97	2.21	
12.09	1.36		23.26	1.79		38.35	2.22	
12.30	1.37		23.58	1.80	36	38.74	2.23	
12.52	1.38		23.89	1.81		39.12	2.24	
12.73	1.39		24.20	1.82		39.51	2.25	45
12.95	1.40	28	24.52	1.83		39.90	2.26	
13.17	1.41		24.84	1.84		40.28	2.27	
13.39	1.42		25.16	1.85	37	40.67	2.28	
13.61	1.43		25.49	1.86		41.06	2.29	
13.84	1.44		25.81	1.87		41.45	2.30	46
14.07	1.45	29	26.14	1.88		41.85	2.31	
14.30	1.46		26.47	1.89		42.24	2.32	
14.53	1.47		26.81	1.90	38	42.63	2.33	
14.77	1.48		27.14	1.91		43.02	2.34	
15.00	1.49		27.48	1.92		43.42	2.35	47
15.25	1.50	30	27.81	1.93		43.81	2.36	
15.49	1.51		28.15	1.94		44.21	2.37	
15.73	1.52		28.50	1.95	39	44.60	2.38	
15.98	1.53		28.84	1.96		45.00	2.39	
16.23	1.54		29.19	1.97		45.40	2.40	48
16.49	1.55	31	29.53	1.98		45.70	2.41	
16.74	1.56		29.88	1.99		46.19	2.42	

TABLE V (*continued*)

Per cent failing	*Distance* $\frac{x}{\sigma}$	*Percentile score*	*Per cent failing*	*Distance* $\frac{x}{\sigma}$	*Percentile score*	*Per cent failing*	*Distance* $\frac{y}{\sigma}$	*Percentile score*
46.59	2.43		64.68	2.86		79.14	3.29	
46.99	2.44		65.05	2.87		79.43	3.30	66
47.39	2.45	49	65.42	2.88		79.72	3.31	
47.79	2.46		65.79	2.89		80.01	3.32	
48.18	2.47		66.16	2.90	58	80.29	3.33	
48.58	2.48		66.53	2.91		80.57	3.34	
48.98	2.49		66.90	2.92		80.85	3.35	67
50.00	2.50	50	67.26	2.93		81.13	3.36	
51.02	2.51		67.62	2.94		81.40	3.37	
51.42	2.52		67.88	2.95	59	81.68	3.38	
51.82	2.53		68.34	2.96		81.95	3.39	
52.21	2.54		68.70	2.97		82.21	3.40	68
52.61	2.55	51	69.06	2.98		82.48	3.41	
53.01	2.56		69.41	2.99		82.74	3.42	
53.41	2.57		69.77	3.00	60	83.00	3.43	
53.81	2.58		70.12	3.01		83.26	3.44	
54.21	2.59		70.47	3.02		83.51	3.45	69
54.60	2.60	52	70.81	3.03		83.77	3.46	
55.00	2.61		71.16	3.04		84.02	3.47	
55.40	2.62		71.50	3.05	61	84.27	3.48	
55.79	2.63		71.85	3.06		84.51	3.49	
56.19	2.64		72.19	3.07		84.75	3.50	70
56.58	2.65	53	72.52	3.08		85.00	3.51	
56.98	2.66		72.86	3.09		85.23	3.52	
57.37	2.67		73.19	3.10	62	85.47	3.53	
57.76	2.68		73.53	3.11		85.70	3.54	
58.15	2.69		73.86	3.12		85.93	3.55	71
58.55	2.70	54	74.19	3.13		86.16	3.56	
58.94	2.71		74.51	3.14		86.39	3.57	
59.33	2.72		74.84	3.15	63	86.61	3.58	
59.72	2.73		75.16	3.16		86.83	3.59	
60.10	2.74		75.48	3.17		87.05	3.60	72
60.49	2.75	55	75.80	3.18		87.27	3.61	
60.88	2.76		76.11	3.19		87.48	3.62	
61.26	2.77		76.42	3.20	64	87.70	3.63	
61.65	2.78		76.74	3.21		87.91	3.64	
62.03	2.79		77.04	3.22		88.11	3.65	73
62.41	2.80	56	77.35	3.23		88.32	3.66	
62.79	2.81		77.66	3.24		88.52	3.67	
63.07	2.82		77.96	3.25	65	88.72	3.68	
63.55	2.83		78.26	3.26		88.92	3.69	
63.93	2.84		78.56	3.27		89.11	3.70	74
64.30	2.85	57	78.85	3.28		89.31	3.71	

TABLE V (*continued*)

Per cent failing	*Distance* $\frac{x}{\sigma}$	*Percentile score*	*Per cent failing*	*Distance* $\frac{x}{\sigma}$	*Percentile score*	*Per cent failing*	*Distance* $\frac{x}{\sigma}$	*Percentile score*
89.50	3.72		95.67	4.15	83	98.74	4.58	
89.69	3.73		95.77	4.16		98.79	4.59	
89.87	3.74		95.87	4.17		98.83	4.60	92
90.06	3.75	75	95.97	4.18		98.88	4.61	
90.24	3.76		96.07	4.19		98.92	4.62	
90.42	3.77		96.16	4.20	84	98.96	4.63	
90.59	3.78		96.26	4.21		99.00	4.64	
90.77	3.79		96.35	4.22		99.04	4.65	93
90.94	3.80	76	96.44	4.23		99.08	4.66	
91.11	3.81		96.53	4.24		99.12	4.67	
91.28	3.82		96.61	4.25	85	99.16	4.68	
91.45	3.83		96.70	4.26		99.23	4.69	
91.61	3.84		96.78	4.27		99.27	4.70	94
91.77	3.85	77	96.87	4.28			4.71	
91.93	3.86		96.95	4.29		99.30	4.72	
92.09	3.87		97.03	4.30	86	99 33	4.73	
92.24	3.88		97.11	4.31		99.37	4.74	
92.39	3.89		97.18	4.32		99.40	4.75	95
92.54	3.90	78	97.26	4.33		99.43	4.76	
92.69	3.91		97.33	4.34		99.46	4.77	
92.84	3.92		97.40	4.35	87	99.49	4.78	
92.98	3.93		97.48	4.36		99.52	4.79	
93.13	3.94		97.55	4.37		99.55	4.80	96
93.27	3.95	79	97.61	4.38		99.58	4.81	
93.41	3.96		97.68	4.39		99.60	4.82	
93.54	3.97		97.75	4.40	88	99.63	4.83	
93.68	3.98		97.81	4.41		99.66	4.84	
93.81	3.99		97.88	4.42		99.68	4.85	97
93.94	4.00	80	97.94	4.43		99.71	4.86	
94.07	4.01		98.00	4.44		99.73	4.87	
94.19	4.02		98.06	4.45	89	99.75	4.88	
94.32	4.03		98.12	4.46		99.78	4.89	
94.44	4.04		98.20	4.47		99.80	4.90	
94.56	4.05	81	98.24	4.48		99.82	4.91	98
94.68	4.06		98.29	4.49		99.84	4 92	
94.80	4.07		98.35	4.50	90	99.87	4.93	
94.92	4.08		98.40	4.51		99.88	4.94	
95.03	4.09		98.45	4.52		99.91	4.95	99
95.14	4.10	82	98.50	4.53		99.93	4.96	
95.25	4.11		98.55	4.54		99.94	4.97	
95.36	4.12		98.60	4.55	91	99.96	4.98	
95.47	4.13		98.65	4.56		99.98	4.99	
95.57	4.14		98.70	4.57		100.00	5.00	100

TABLE VI

Percentile scores to be assigned to test problems or questions which correspond to various percentages of pupils who fail to solve problems or questions correctly. Table is based upon area of probability curve, assuming base line to be broken off at $\pm 3.0\sigma$. Scholastic abilities are assumed to fit the probability curve and percentages of pupils who solve various problems correspond to percentages of area under the curve from the 0 point to a point on the base line. This point on base line, measured in units of σ, is transformed into percentile scores by setting 0 at $-3.0\,\sigma$, 50 at the mean and 100 at $+3.0\sigma$.

Example: A problem failed by 3 per cent of a large number of pupils is scored 35; one failed by 80.09 per cent is scored 64, etc.

Per cent failing	Distance $\frac{x}{\sigma}$	Percentile score	Per cent failing	Distance $\frac{x}{\sigma}$	Percentile score	Per cent failing	Distance $\frac{x}{\sigma}$	Percentile score	Per cent failing	Distance $\frac{x}{\sigma}$	Percentile score
.00	.01		.19	.28		.57	.55		1.32	.82	
.00	.02		.20	.29		.59	.56		1.36	.83	
.01	.03		.21	.30	5	.61	.57		1.40	.84	14
.01	.04		.22	.31		.64	.58		1.44	.85	
.02	.05		.23	.32		.66	.59		1.48	.86	
.02	.06	1	.24	.33		.68	.60	10	1.52	.87	
.03	.07		.25	.34		.70	.61		1.56	.88	
.04	.08		.26	.35		.73	.62		1.60	.89	
.04	.09		.27	.36	6	.75	.63		1.65	.90	15
.05	.10		.29	.37		.77	.64		1.69	.91	
.05	.11		.30	.38		.80	.65		1.74	.92	
.06	.12	2	.31	.39		.82	.66	11	1.78	.93	
.07	.13		.33	.40		.85	.67		1.83	.94	
.07	.14		.34	.41		.88	.68		1.88	.95	
.08	.15		.35	.42	7	.90	.69		1.93	.96	16
.09	.16		.37	.43		.93	.70		1.98	.97	
.09	.17		.38	.44		.96	.71		2.03	.98	
.10	.18	3	.40	.45		.99	.72	12	2.08	.99	
.11	.19		.41	.46		1.02	.73		2.14	1.00	
.12	.20		.43	.47		1.05	.74		2.19	1.01	
.12	.21		.45	.48	8	1.08	.75		2.25	1.02	17
.13	.22		.46	.49		1.11	.76		2.30	1.03	
.14	.23		.48	.50		1.15	.77		2.36	1.04	
.15	.24	4	.50	.51		1.18	.78	13	2.42	1.05	
.16	.25		.52	.52		1.22	.79		2.48	1.06	
.17	.26		.54	.53		1.25	.80		2.54	1.07	
.18	.27		.55	.54	9	1.29	.81		2.60	1.08	18

TABLE VI (*continued*)

Per cent failing	*Distance* $\frac{x}{\sigma}$	*Percentile score*	*Per cent failing*	*Distance* $\frac{x}{\sigma}$	*Percentile score*	*Per cent failing*	*Distance* $\frac{x}{\sigma}$	*Percentile score*
2.67	1.09		6.67	1.51		14.09	1.93	
2.73	1.10		6.80	1.52		14.32	1.94	
2.80	1.11		6.94	1.53		14.55	1.95	
2.87	1.12		7.07	1.54		14.78	1.96	
2.93	1.13		7.21	1.55		15.01	1.97	
3.00	1.14	19	7.35	1.56	26	15.25	1.98	33
3.08	1.15		7.50	1.57		15.48	1.99	
3.15	1.16		7.64	1.58		15.73	2.00	
3.22	1.17		7.79	1.59		15.97	2.01	
3.30	1.18		7.94	1.60		16.21	2.02	
3.37	1.19		8.09	1.61		16.46	2.03	
3.45	1.20	20	8.24	1.62	27	16.71	2.04	34
3.53	1.21		8.39	1.63		16.96	2.05	
3.61	1.22		8.55	1.64		17.22	2.06	
3.70	1.23		8.71	1.65		17.48	2.07	
3.78	1.24		8.87	1.66		17.74	2.08	
3.87	1.25		9.04	1.67		18.00	2.09	
3.95	1.26	21	9.20	1.68	28	18.27	2.10	35
4.04	1.27		9.37	1.69		18.53	2.11	
4.13	1.38		9.54	1.70		18.80	2.12	
4.22	1.29		9.71	1.71		19.08	2.13	
4.32	1.30		9.89	1.72		19.35	2.14	
4.41	1.31		10.06	1.73		19.63	2.15	
4.51	1.32	22	10.24	1.74	29	19.91	2.16	36
4.61	1.33		10.42	1.75		20.19	2.17	
4.71	1.34		10.61	1.76		20.47	2.18	
4.81	1.35		10.79	1.77		20.76	2.19	
4.91	1.36		10.98	1.78		21.05	2.20	
5.02	1.37		11.17	1.79		21.34	2.21	
5.12	1.38	23	11.37	1.80	30	21.63	2.22	37
5.23	1.39		11.56	1.81		21.92	2.23	
5.34	1.40		11.76	1.82		22.22	2.24	
5.45	1.41		11.96	1.83		22.52	2.25	
5.57	1.42		12.16	1.84		22.82	2.26	
5.68	1.43		12.37	1.85		23.13	2.27	
5.80	1.44	24	12.57	1.86	31	23.44	2.28	38
5.92	1.45		12.78	1.87		23.75	2.29	
6.03	1.46		13.00	1.88		24.06	2.30	
6.16	1.47		13.21	1.89		24.32	2.31	
6.29	1.48		13.43	1.90		24.69	2.32	
6.41	1.49		13.65	1.91		25.00	2.33	
6.54	1.50	25	13.87	1.92	32	25.32	2.34	39

TABLE VI (*continued*)

Per cent failing	*Distance* $\frac{x}{\sigma}$	*Percentile score*	*Per cent failing*	*Distance* $\frac{x}{\sigma}$	*Percentile score*	*Per cent failing*	*Distance* $\frac{x}{\sigma}$	*Percentile score*
25.64	2.35		40.76	2.77		57.67	3.19	
25.97	2.36		41.15	2.78		58.07	3.20	
26.29	2.37		41.54	2.79		58.46	3.21	
26.62	2.38		41.93	2.80		58.85	3.22	
26.95	2.39		42.33	2.81		59.24	3.23	
27.29	2.40	40	42.72	2.82	47	59.62	3.24	54
27.62	2.41		43.11	2.83		60.01	3.25	
27.96	2.42		43.50	2.84		60.40	3.26	
28.29	2.43		43.90	2.85		60.78	3.27	
28.63	2.44		44.29	2.86		61.17	3.28	
28.98	2.45		44.69	2.87		61.55	3.29	
29.32	2.46	41	45.08	2.88	48	61.93	3.30	55
29.67	2.47		45.48	2.89		62.31	3.31	
30.01	2.48		45.88	2.90		62.69	3.32	
30.36	2.49		46.27	2.91		63.07	3.33	
30.71	2.50		46.67	2.92		63.45	3.34	
31.07	2.51		47.07	2.93		63.82	3.35	
31.42	2.52	42	47.47	2.94	49	64.20	3.36	56
31.78	2.53		47.87	2.95		64.57	3.37	
32.14	2.54		48.26	2.96		64.94	3.38	
32.50	2.55		48.66	2.97		65.31	3.39	
32.86	2.56		49.06	2.98		65.68	3.40	
33.22	2.57		49.46	2.99		66.05	3.41	
33.58	2.58	43	50.00	3.00	50	66.42	3.42	57
33.95	2.59		50.54	3.01		66.78	3.43	
34.32	2.60		50.94	3.02		67.14	3.44	
34.69	2.61		51.34	3.03		67.50	3.45	
35.06	2.62		51.74	3.04		67.86	3.46	
35.43	2.63		52.13	3.05		68.22	3.47	
35.80	2.64	44	52.53	3.06	51	68.58	3.48	58
36.18	2.65		52.93	3.07		68.93	3.49	
36.55	2.66		53.33	3.08		69.29	3.50	
36.93	2.67		53.73	3.09		69.64	3.51	
37.31	2.68		54.12	3.10		69.99	3.52	
37.69	2.69		54.52	3.11		70.33	3.53	
38.07	2.70	45	54.92	3.12	52	70.63	3.54	59
38.45	2.71		55.31	3.13		71.02	3.55	
38.83	2.72		55.71	3.14		71.37	3.56	
39.22	2.73		56.10	3.15		71.71	3.57	
39.60	2.74		56.50	3.16		72.04	3.58	
39.99	2.75		56.89	3.17		72.38	3.59	
40.38	2.76	46	57.28	3.18	53	72.71	3.60	60

TABLE VI (*continued*)

Per cent failing	*Distance* $\frac{x}{\sigma}$	*Percentile score*	*Per cent failing*	*Distance* $\frac{x}{\sigma}$	*Percentile score*	*Per cent failing*	*Distance* $\frac{x}{\sigma}$	*Percentile score*
73.05	3.61		84.99	4.03		92.79	4.45	
73.38	3.62		85.22	4.04		92.93	4.46	
73.71	3.63		85.45	4.05		93.06	4.47	
74.03	3.64		85.68	4.06		93.20	4.48	
74.36	3.65		85.91	4.07		93.33	4.49	
74.68	3.66	61	86.13	4.08	68	93.46	4.50	75
75.00	3.67		86.35	4.09		93.59	4.51	
75.31	3.68		86.57	4.10		93.71	4.52	
75.63	3.69		86.79	4.11		93.84	4.53	
75.94	3.70		87.00	4.12		93.97	4.54	
76.25	3.71		87.22	4.13		94.08	4.55	
76.56	3.72	62	87.43	4.14	69	94.20	4.56	76
76.87	3.73		87.63	4.15		94.32	4.57	
77.18	3.74		87.84	4.16		94.43	4.58	
77.48	3.75		88.04	4.17		94.55	4.59	
77.78	3.76		88.24	4.18		94.66	4.60	
78.08	3.77		88.44	4.19		94.77	4.61	
78.37	3.78	63	88.63	4.20	70	94.88	4.62	77
78.66	3.79		88.83	4.21		94.98	4.63	
78.95	3.80		89.02	4.22		95.09	4.64	
79.24	3.81		89.21	4.23		95.19	4.65	
79.53	3.82		89.39	4.24		95.29	4.66	
79.81	3.83		89.58	4.25		95.39	4.67	
80.09	3.84	64	89.76	4.26	71	95.49	4.68	78
80.37	3.85		89.94	4.27		95.59	4.69	
80.65	3.86		90.11	4.28		95.68	4.70	
80.92	3.87		90.29	4.29		95.78	4.71	
81.20	3.88		90.46	4.30		95.87	4.72	
81.47	3.89		90.63	4.31		95.96	4.73	
81.73	3.90	65	90.80	4.32	72	96.05	4.74	79
82.00	3.91		90.96	4.33		96.13	4.75	
82.26	3.92		91.13	4.34		96.22	4.76	
82.52	3.93		91.29	4.35		96.30	4.77	
82.78	3.94		91.45	4.36		96.39	4.78	
83.04	3.95		91.61	4.37		96.47	4.79	
83.29	3.96	66	91.76	4.38	73	96.55	4.80	80
83.54	3.97		91.91	4.39		96.63	4.81	
83.79	3.98		92.06	4.40		96.70	4.82	
84.03	3.99		92.21	4.21		96.78	4.83	
84.27	4.00		92.36	4.42		96.85	4.84	
84.52	4.01		92.50	4.43		96.92	4.85	
84.75	4.02	67	92.65	4.44	74	97.00	4.86	81

TABLE VI (*continued*)

Per cent failing	*Distance* $\frac{x}{\sigma}$	*Percentile score*	*Per cent failing*	*Distance* $\frac{x}{\sigma}$	*Percentile score*	*Per cent failing*	*Distance* $\frac{x}{\sigma}$	*Percentile score*
97.00	4.87		98.92	5.25		99.71	5.63	
97.13	4.88		98.95	5.26		99.73	5.64	94
97.20	4.89		98.98	5.27		99.74	5.65	
97.27	4.90		99.01	5.28	88	99.75	5.66	
97.33	4.91		99.04	5.29		99.76	5.67	
97.40	4.92	82	99.07	5.30		99.77	5.68	
97.46	4.93		99.10	5.31		99.78	5.69	
97.52	4.94		99.12	5.32		99.79	5.70	95
97.58	4.95		99.15	5.33		99.80	5.71	
97.64	4.96		99.18	5.34	89	99.81	5.72	
97.70	4.97		99.20	5.35		99.82	5.73	
97.75	4.98	83	99.23	5.36		99.83	5.74	
97.81	4.99		99.25	5.37		99.84	5.75	
97.86	5.00		99.27	5.38		99.85	5.76	96
97.92	5.01		99.30	5.39		99.86	5.77	
97.97	5.02		99.32	5.40	90	99.87	5.78	
98.02	5.03		99.34	5.41		99.88	5.79	
98.07	5.04	84	99.36	5.42		99.88	5.80	
98.12	5.05		99.39	5.43		99.89	5.81	
98.17	5.06		99.41	5.44		99.90	5.82	97
98.22	5.07		99.43	5.45		99.91	5.83	
98.26	5.08		99.45	5.46	91	99.91	5.84	
98.31	5.09		99.46	5.47		99.92	5.85	
98.35	5.10	85	99.48	5.45		99.93	5.86	
98.40	5.11		99.50	5.49		99.93	5.87	
98.44	5.12		99.52	5.50		99.94	5.88	98
98.48	5.13		99.54	5.51		99.95	5.89	
98.52	5.14		99.55	5.52	92	99.95	5.90	
98.56	5.15		99.57	5.53		99.96	5.91	
98.60	5.16	86	99.59	5.54		99.96	5.92	
98.64	5.17		99.60	5.55		99.97	5.93	
98.68	5.18		99.62	5.56		99.98	5.94	99
98.71	5.19		99.63	5.57		99.98	5.95	
98.75	5.20		99.65	5.58	93	99.99	5.96	
98.78	5.21		99.66	5.59		99.99	5.97	
98.82	5.22	87	99.67	5.60		100.00	5.98	
98.85	5.23		99.69	5.61		100.00	5.99	
98.89	5.24		99.70	5.62		100.00	6.00	100

TABLE VII

Values of r for corresponding values of ρ. ρ is computed from the expression, $\rho = 1 - \frac{6 \Sigma D^2}{N(N^2 - 1)}$

r could be computed from, $r = 2 \sin \left(\frac{\pi}{6}\rho\right)$.

Values of r given in this table have been computed for various values of ρ ranging from .01 to 1.00.

ρ	r	ρ	r	ρ	r	ρ	r
.01	.0105	.26	.2714	.51	.5277	.76	.7750
.02	.0209	.27	.2818	.52	.5378	.77	.7847
.03	.0314	.28	.2922	.53	.5479	.78	.7943
.04	.0419	.29	.3025	.54	.5580	.79	.8039
.05	.0524	.30	.3129	.55	.5680	.80	.8135
.06	.0628	.31	.3232	.56	.5781	.81	.8230
.07	.0733	.32	.3335	.57	.5881	.82	.8325
.08	.0838	.33	.3439	.58	.5981	.83	.8421
.09	.0942	.34	.3542	.59	.6081	.84	.8516
.10	.1047	.35	.3645	.60	.6180	.85	.8610
.11	.1151.	.36	.3748	.61	.6280	.86	.8705
.12	.1256	.37	.3850	.62	.6379	.87	.8799
.13	.1360	.38	.3935	.63	.6478	.88	.8893
.14	.1465	.39	.4056	.64	.6577	.89	.8986
.15	.1569	.40	.4158	.65	.6676	.90	.9080
.16	.1674	.41	.4261	.66	.6775	.91	.9173
.17	.1778	.42	.4363	.67	.6873	.92	.9269
.18	.1882	.43	.4465	.68	.6971	.93	.9359
.19	.1986	.44	.4567	.69	.7069	.94	.9451
.20	.2091	.45	.4669	.70	.7167	.95	.9543
.21	.2195	.46	.4771	.71	.7265	.96	.9635
.22	.2299	.47	.4872	.72	.7363	.97	.9727
.23	.2403	.48	.4973	.73	.7460	.98	.9818
.24	.2507	.49	.5075	.74	.7557	.99	.9909
.25	.2611	.50	.5176	.75	.7654	1.00	1.0000

TABLE VIII

Values of r for corresponding values of R, R having been computed from

$$R = 1 - \frac{6\,\Sigma\,G}{N^2 - 1}$$

r could be computed from the expression

$$r = 2\cos\frac{\pi}{3}(1-R)-1.$$

Values of r given in this table have been computed for values of R ranging from .01 to 1.00.

R	r	R	r	R	r	R	r
.00	.000	.26	.429	.51	.742	.76	.937
.01	.018	.27	.444	.52	.753	.77	.942
.02	.036	.28	.458	.53	.763	.78	.947
.03	.054	.29	.472	.54	.772	.79	.952
.04	.071	.30	.486	.55	.782	.80	.956
.05	.089	.31	.500	.56	.791	.81	.961
.06	.107	.32	.514	.57	.801	.82	.965
.07	.124	.33	.528	.58	.810	.83	.968
.08	.141	.34	.541	.59	.818	.84	.972
.09	.158	.35	.554	.60	.827	.85	.975
.10	.176	.36	.567	.61	.836	.86	.979
.11	.192	.37	.580	.62	.844	.87	.981
.12	.209	.38	.593	.63	.852	.88	.984
.13	.226	.39	.606	.64	.860	.89	.987
.14	.242	.40	.618	.65	.687	.90	.989
.15	.259	.41	.630	.66	.875	.91	.991
.16	.275	.42	.642	.67	.882	.92	.993
.17	.291	.43	.654	.68	.889	.93	.995
.18	.307	.44	.666	.69	.896	.94	.996
.19	.323	.45	.677	.70	.902	.95	.997
.20	.338	.46	.689	.71	.908	.96	.998
.21	.354	.47	.700	.72	.915	.97	.999
.22	.369	.48	.711	.73	.921	.98	.9996
.23	.384	.49	.721	.74	.926	.99	.9999
.24	.399	.50	.732	.75	.932	1.00	1.0000
.25	.414						

Table IX

Values of *r* corresponding to various percentages of unlike-signed pairs. *U* represents the percentage that the number of pairs of measures having "unlike signs" (*i.e.*, the number of pairs in which each member is above the mean in one series and below the mean in the other series) is of the total number of pairs.

U	*r*	*U*	*r*	*U*	*r*	*U*	*r*
.00	1.0000	.13	.9174	.26	.6848	.38	.3682
.01	.9996	.14	.9044	.27	.6615	.39	.3387
.02	.9982	.15	.8905	.28	.6375	.40	.3089
.03	.9958	.16	.8757	.29	.6129	.41	.2788
.04	.9924	.17	.8602	.30	.5877	.42	.2485
.05	.9880	.18	.8439	.31	.5620	.43	.2180
.06	.9826	.19	.8268	.32	.5358	.44	.1873
.07	.9762	.20	.8089	.33	.5091	.45	.1564
.08	.9688	.21	.7902	.34	.4819	.46	.1253
.09	.9604	.22	.7707	.35	.4542	.47	.0941
.10	.9510	.23	.7504	.36	.4260	.48	.0628
.11	.9407	.24	.7293	.37	.3973	.49	.0314
.12	.9295	.25	.7074			.50	.0000

TABLE X

PROBABLE ERRORS OF THE COEFFICIENT OR CORRELATION FOR VARIOUS NUMBERS OF MEASURES (N) AND FOR VARIOUS VALUES OF r.

Number of Measures	*Correlation Coefficient r.*						
	0.0	*0.1*	*0.2*	*0.3*	*0.4*	*0.5*	*0.6*
20	1508	1493	1448	1373	1267	1131	0965
30	1231	1219	1182	1121	1035	0924	0788
40	1067	1056	1024	0971	0896	0800	0683
50	0954	0944	0915	0868	0801	0715	0610
70	0806	0798	0774	0734	0677	0605	0516
100	0674	0668	0648	0614	0567	0506	0432
150	0551	0546	0529	0501	0463	0413	0352
200	0477	0472	0458	0434	0401	0358	0305
250	0426	0421	0409	0387	0358	0319	0272
300	0389	0386	0374	0354	0327	0292	0249
400	0337	0334	0324	0307	0283	0253	0216
500	0302	0299	0290	0274	0253	0226	0193
1000	.0213	0211	0205	0194	0179	0160	0137

Number of Measures	*0.65*	*0.7*	*0.75*	*0.8*	*0.85*	*0.9*	*0.95*
20	0871	0769	0660	0543	0419	0287	0147
30	0711	0628	0539	0444	0342	0234	0120
40	0616	0544	0467	0384	0296	0203	0104
50	0551	0486	0417	0343	0265	0181	0093
70	0466	0411	0353	0290	0224	0153	0079
100	0391	0345	0294	0242	0187	0128	0066
150	0318	0281	0241	0198	0153	0105	0054
200	0275	0243	0209	0172	0133	0091	0047
250	0246	0218	0187	0154	0118	0081	0042
300	0225	0199	0170	0140	0108	0074	0038
400	0195	0172	0148	0122	0094	0064	0033
500	0174	0154	0132	0109	0084	0057	0029
1000	0123	0109	0093	0077	0059	0041	0021

INDEX

harge, book must be brought to the desk.

TWO WEEK BOOK

DO NOT RETURN BOOKS ON SUNDAY

DATE DUE

~~MY 21 1953~~

~~[illegible]~~

PD

~~4 12-Ap '16~~

Ⓝ

~~453 JUN 4~~